UNIVERSITY OF HAWAII
LIBRARY
Nov 29 '50

The PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

. 41 SEVENTH SERIES No. 322 November, 1950

A Journal of Theoretical Experimental and Applied Physics

EDITOR
PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD
SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.
ALLAN FERGUSON, M.A., D.Sc.
SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.
PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

PRICE 10s.
Annual Subscription £5 2s. 6d. payable in advance.

ED AND PUBLISHED BY TAYLOR & FRANCIS LTD., RED LION COURT, FLEET ST., LONDON, E.C.4.

Annals of Science

A QUARTERLY REVIEW OF
THE HISTORY OF SCIENCE
SINCE THE RENAISSANCE

EDITORS

D. McKIE, D.Sc., Ph.D.,
University College, London.

HARCOURT BROWN, M.A., Ph.D.,
Brown University, Providence, R.I., U.S.A.

H. W. ROBINSON,
Former Librarian,
Royal Society of London.

ANNUAL SUBSCRIPTION
£2 0s. 0d.
OR
10s. 6d.
PER PART
POST FREE

TAYLOR & FRANCIS, LTD., Red Lion Court, Fleet Street, LONDON, E.C.4

ALERE FLAMMAM

THE MATHEMATICAL WORKS OF JOHN WALLIS, D.D., F.R.S.

by

J. F. SCOTT, Ph.D., B.A.

"His work will be indispensable to those interested in the early history of The Royal Society. I commend to all students of the Seventeenth Century, whether scientific or humane, this learned and lucid book."—Extract from foreword by Prof. E. N. da C. Andrade, D.Sc., Ph.D., F.R.S.

Recommended for publication by University of London

12/6 net

Printed and Published by:

TAYLOR & FRANCIS, LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4.

CONTENTS OF No. 322.

Page

XCVII. Deformation Textures in Face-Centred Cubic Metals. By E. A. Calnan, B.Sc., and C. J. B. Clews, Ph.D., National Physical Laboratory, Metallurgy Division 1085

XCVIII. The Energy Distribution behind Decaying Shocks.—I. Plane Waves. By M. J. Lighthill.............................. 1101

XCIX. Energy Losses of Ionizing Particles at Relativistic Velocities in the Photographic Plate. By H. Messel and D. M. Ritson, Dublin Institute for Advanced Studies 1129

C. A Theory of the Plastic Bulging of a Metal Diaphragm by Lateral Pressure. By R. Hill, M.A., Ph.D., Metal Flow Research Laboratory, Sheffield 1133

CI. A Method for the Mapping of Vector Potential Distributions in Axially Symmetrical Systems. By G. Liebmann, Associated Electrical Industries Research Laboratory, Aldermaston, Berks 1143

CII. Double Stars due to Cosmic Rays. By T. T. Li, Imperial College, London 1152

CIII. Communication Theory and Physics. By D. Gabor, Imperial College, London 1161

CIV. Density Variations in Aluminium. By J. L. Snoek, Philips Research Laboratories, Eindhoven 1188

CV. Correspondence :—
  Radiative Corrections in the Non-Relativistic Limit. By W. Thirring, D.Phil., Dublin Institute for Advanced Studies. 1193
  Multiple Processes in the Production of Pairs of Electrons by γ-Radiation. By J. E. Hooper and D. T. King, H. H. Wills Physical Laboratory. (Plate XXXV.) 1194

CVI. Notices of New Books and Periodicals received :—
  H. Dingle's Practical Applications of Spectrum Analysis .. 1196
  D. H. Bangham's Progress in Coal Science 1197
  Martin Johnson's Astronomy of Stellar Energy and Decay.. 1197
  D. H. Bangham's Progress in Coal Science 1197

***All communications for the Philosophical Magazine should be addressed, post-paid, to the Editors, c/o Messrs. Taylor and Francis, Ltd., Red Lion Court, Fleet Street, London, England.

XCVII. *Deformation Textures in Face-Centred Cubic Metals.*

By E. A. Calnan, B.Sc., and C. J. B. Clews, Ph.D.
National Physical Laboratory, Metallurgy Division*.

[Received July 11, Revised August 8, 1950.]

Introduction.

There have been several previous attempts to explain the deformation behaviour of polycrystalline aggregates in terms of the slip mechanisms operative when single crystals are deformed, but no wholly satisfactory treatment has been devised. In discussing these, reference must first be made to the work of Boas and Schmid (1931), which assumes the operation of the three most favourable slip systems for any particular crystal orientation and derives the end position which is stable for these three systems. This predicts tension and compression textures, for both face-centred and body-centred cubic metals, which are in good agreement with the main features of the observed textures. Some details of the observed textures are, however, not explained. The treatment of Taylor (1938), mathematically more rigorous, has as its basic assumption the perfectly homogeneous deformation of the aggregate. Following the analysis of Mises (1928) that this demands the simultaneous operation of five special glide systems, Taylor has derived the systems and corresponding shears which conform to the principle of virtual work. The magnitudes and directions of the crystal rotations may be deduced from the shears selected, while the critical shear stress for a particular elongation is calculated from the sum of the shears and the single crystal shear-hardening curve. Good agreement with the experimentally determined tensile stress-elongation curve for a polycrystalline aggregate, and the main features of the tension and compression textures for face-centred cubic metals are obtained. The disadvantages of the treatment are first that homogeneous deformation rarely occurs in practice, and secondly that the heavy computation makes its application to metals of other crystal structures virtually impossible.

Pickus and Mathewson (1939) have put forward a theory of the origin of rolling textures in which they assume that unequal participation of the three most favourable glide systems brings the deformation axes to symmetrical end positions, where the resolved shear stresses are equal and the rotations cancel out. The most favourable systems are those for which the product of the resolved shear stress and the cosine of the angle between the glide direction and the direction in which free flow may occur are highest. The relative probability of the end positions are considered to depend on the values of the product function at the end positions. It is shown that all the observed rolling textures in face-centred metals fit in reasonably well with these ideas. The main criticism of this treatment and that of Boas and Schmid appears to be that, although the crystals oriented to correspond to the end positions will no doubt remain there and will slip without rotation, there is no complete explanation of how other crystals reach these positions.

---

* Communicated by the Authors.

One further criticism of all the previous treatments may be noted. Although the simultaneous operation of several slip systems is demanded, there is no explanation of the means by which the most favourable system or systems can apparently withstand resolved shear stresses greater than the critical value until the resolved shear stresses on less favourable systems have reached the critical value and deformation occurs.

MAIN CONCEPTS OF THE PRESENT TREATMENT.

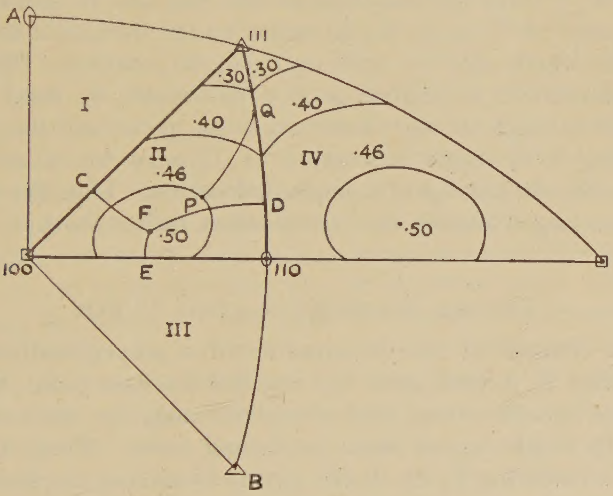
It is an experimental fact that considerable deformation can occur without cracks appearing at the grain boundaries. In the ideal case of homogeneous deformation this demands the operation of five special slip systems, and even allowing for some inhomogeneous deformation, particularly near the grain boundaries, the critical shear stresses must be attained on more than one system at the same time. Moreover, apart from minor differences, for example, the observations on duplex slip in α-brass crystals, Göler and Sachs (1929), the critical shear stresses must be equal. Therefore the constraints imposed by the surrounding grains of the aggregate together with the applied stress must be such as to produce equal resolved shear stresses on the glide systems which operate. That is to say, if a grain deforms by the simultaneous operation of two systems then the resolved shear stress on these systems must be equal, and consequently it will be assumed that the effective stress may be represented by a tensile stress, T_e, so oriented as to give duplex slip, *i.e.* its representative point on the stereographic projection lies on one of the boundaries of the unit triangle. If, however, the grain deforms by the operation of more than two slip systems then T_e must lie at one of the corner points of the triangle, [100], [110], and [111], where there are respectively 8, 4, and 6 equally favourable slip systems.

This concept, although probably an over-simplification of what must be a complex system of stresses, does give a physical picture of the mechanism. Further consideration of the movement of T_e leads to an explanation of the means by which T_e reaches positions favourable for duplex or multiple slip without the resolved shear stress on the most favourable system exceeding the critical value.

The most favourable system for a grain of applied tension axis in unit triangle II of fig. 1 is the direction A on the slip plane of normal B. As the applied stress, T_a, rises the resolved shear stress rises proportionately until it reaches the critical value at which slip would occur if the grain were an isolated single crystal. In general this single slip mechanism does not operate for all the grains in the polycrystalline aggregate, and therefore the constraints due to the surrounding grains must prevent the resolved shear stress from reaching the critical value. Due to these constraints the direction of the effective stress, T_e, moves away from the direction of the applied stress in such a way that T_e resolved on the most favourable system is always below the critical value.

In fig. 1 are shown the contours of equal resolved shear stress for the most favourable system, the numerical values being $\cos\chi \cos\lambda$, where χ and λ are respectively the angles which the slip direction and slip plane normal make with the particular T_a orientation. For orientation P (fig. 1), the critical resolved shear stress C is reached when $T_a = C/0{\cdot}46$, and in order that the critical value shall not be reached as T_a increases, T_e will move away to a point of lower $\cos\chi \cos\lambda$ value, *i.e.*, move down the contour gradient. When T_a has increased such that $T_a = C/0{\cdot}40$, and slip has not occurred, then T_e will lie on the 0·40 contour. The position on the contour will be determined by the stresses exerted by the neighbouring grains, and with a fine grained polycrystalline aggregate the general direction of movement may be taken as the path of quickest descent, *i.e.* perpendicular to the contours. On reaching a boundary of the unit triangle, T_e will travel along it in the direction of decreasing $\cos\chi \cos\lambda$ until it reaches one of the corner points. These are minimum positions, and any further increase in T_a must now result in slip, for in whatever direction T_e attempts to move from the corner point some system will

Fig. 1.



Stereographic projection showing shear stress contours for a face-centred cubic metal.

attain the critical shear stress. Here, then, with the multiple systems available, deformation can occur without immediate misfit at the grain boundaries, and with the symmetrically disposed systems all exactly equivalent there will be no resultant rotation of the crystal.

In some grains T_e may not reach the corner point before slip occurs owing to insufficient lateral stresses, as perhaps in the case of a grain at a free surface. Consequently, rotation will occur in the usual directions of single or duplex slip, the latter taking place if T_e has reached the unit triangle boundary. In addition to those at the surface, the grains most likely to slip before T_e reaches a corner point are those for which the greatest lateral stresses will be required. These, which are the grains most likely to rotate, will thus be those furthest removed from the appropriate corner point, and conversely rotation will be least probable near the corners. In the instant at which slip occurs the lateral stresses are released and T_e returns to T_a or, if rotation takes place, to the new position of T_a. With the cessation of this slip due to strain hardening, the movement of T_e may begin again under the influence of a new environment which may or may not lead to rotation. Since inhomogeneous deformation is implied in this treatment, T_e need not have a unique value throughout the whole of a single grain, and the deformation corresponding to a single movement of T_e may be as small as that associated with the passage of a single dislocation. Thus the deformation process is envisaged as step-like in character rather than as smooth and continuous.

TENSILE STRESS-ELONGATION CURVE.

From the concept of the deformation of a polycrystalline aggregate occurring when T_e in each grain has reached a corner point, the relationship between tensile stress and elongation may be derived from the corresponding single crystal shear hardening curve. First, however, the validity of considering T_e in all the grains as having reached the corner points must be examined, since in this treatment the rotations which lead to deformation textures occur in precisely those grains, or parts of grains, for which this assumption is not true. The fact that the grains do not pull apart at the boundaries must imply that in the boundary regions at least multiple slip occurs. Recent X-ray work on creep deformation in aluminium, Calnan and Burns (1950) has shown the greater complexity of the deformation processes in the regions adjacent to the grain boundaries. In view of the comparatively large size of these boundary regions, in which the lateral stresses must be sufficient to cause considerable displacement of T_e, the lateral stresses elsewhere in the grain are probably not very much less, with the result that T_e will be somewhere near a corner position. It can be shown that the tensile stress for a given elongation is not greatly different for multiple slip deformation (T_e at a corner point) and simple slip deformation with T_e about 5° from the same corner point. Thus an approximation to the polycrystalline curve, within the limits of experimental variations, may be made on the assumption that in all the grains T_e has reached a corner point without vitiating the arguments employed in deriving the deformation textures, namely that in the interior of the grains at least T_e has not reached this point.

On the shear stress contour diagram, fig. 1, the peak position F is joined to the points of maximum shear stress on the boundaries of the unit triangle, C, D and E, by the lines of quickest descent. Since the mean path of T_e is along the line of quickest descent, T_e for all the grains of orientations contained in the area a_{111} between CFD and the [111] moves towards this last direction. For example, T_e for the grain of orientation represented by P moves along its appropriate line for quickest descent until it reaches the [111]–[110] boundary at Q. From the shear stress contours in the adjacent triangle IV, corresponding to a different most favourable slip system, it will be seen that the boundary itself constitutes a line of quickest descent, and thus T_e moves from Q towards [111]. Similarly, T_e for the orientations represented by the area a_{110} between DFE and the [110] moves to the [110], and for those in area a_{100} between CFE and the [100] to the [100]. With an initially random orientation in the polycrystalline aggregate the proportions of grains with these end positions for T_e are given by the ratios of the solid angles to which these areas correspond. It is however more convenient to use the ratios of the areas on the unit triangle, and since this involves an error of only about 1 per cent this method has been adopted.

An elongation e in the direction of the applied stress requires an elongation $e/[\cos^2 \phi(1+\sigma)-\sigma]$ in the direction of a given corner point, where ϕ is the angle between the direction of the applied stress and the corner point, and σ is the value of Poisson's ratio chosen to give approximately constant volume over the range of deformation considered. A weighted mean value of $\cos^2 \phi$ was calculated for each of the three areas referred to in the previous paragraph, viz. $(\cos^2 \phi)_{111}$, $(\cos^2 \phi)_{110}$, and $\cos^2 \phi)_{100}$. For symmetrical slip it can be shown that the elongation in the direction of the corner point, *i.e.* in the direction of the effective stress for multiple slip, is $s \cos \chi_0 \cos \lambda_0$, where s is the sum of the moduli of the shears on the individual operative systems, all of which have the same values of χ_0 and λ_0, the initial angles between the effective stress direction and the slip direction and the slip plane normal respectively. Substituting the appropriate values of χ_0, λ_0 and $(\cos^2 \phi)$ in the expression

$$e=s \cos \chi_0 \cos \lambda_0 [(\cos^2 \phi)(1+\sigma)-\sigma],$$

gives the three values of shear s corresponding to the selected value of e. The critical shear stress for each of these s values may be determined from the experimental shear stress—shear strain curve for a single crystal. The effective tensile stress in the direction of the corner point to produce this shear stress is T_e=critical shear stress/$\cos \chi_0 \cos \lambda_0$, which in turn implies a tensile stress in the applied stress direction

$$T_a=(\cos \phi) \text{ . critical shear stress}/\cos \chi_0 \cos \lambda_0,$$

where $(\cos \phi)$ is the weighted mean value of $\cos \phi$ over the area considered.

TABLE I.

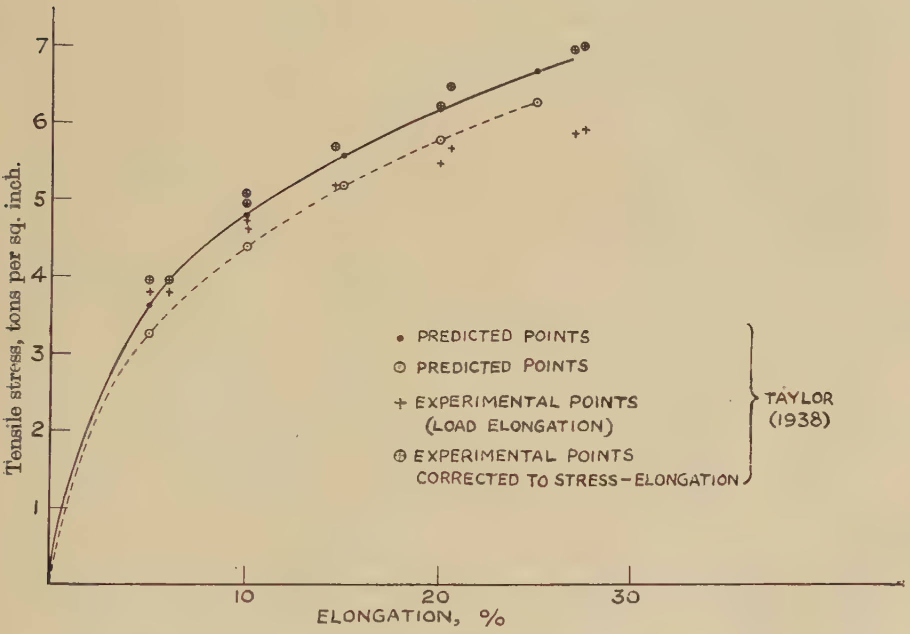
Calculation of the Tensile–Stress–Elongation Curve.

Elongation $e\%$	Weighted mean values $(\cos\phi)_{100}=0·96$ $(\cos^2\phi)_{100}=0·94$			Weighted mean values $(\cos\phi)_{110}=0·97$ $(\cos^2\phi)_{110}=0·95$			Weighted mean values $(\cos\phi)_{111}=0·92$ $(\cos^2\phi)_{111}=0·85$			Weighting factors (100) 0·172	(110) 0·203	(111) 0·625	Total $T_{a_{tsi}}$
	$e'_{100}\%$	$T_{e_{100}tsi}$	$T_{a_{100}tsi}$	$e'_{110}\%$	$T_{e_{110}tsi}$	$T_{a_{110}tsi}$	$e'_{111}\%$	$T_{e_{111}tsi}$	$T_{a_{111}tsi}$	$T_{a_{100}tsi}$	$T_{a_{110}tsi}$	$T_{a_{111}tsi}$	
5	5·5	2·50	2·41	5·4	2·47	2·41	6·4	4·74	4·36	0·42	0·49	2·72	3·62
10	11·0	3·35	3·23	10·7	3·30	3·22	12·8	6·24	5·74	0·56	0·65	3·59	4·80
15	16·5	3·92	3·77	16·1	3·88	3·78	19·2	7·24	6·66	0·65	0·77	4·16	5·58
20	22·0	4·36	4·20	21·5	4·34	4·23	25·6	8·03	7·39	0·72	0·86	4·62	6·20
25	27·5	4·73	4·56	26·9	4·70	4·58	32·0	8·70	8·00	0·78	0·93	5·00	6·71

$e'=e/(\cos^2\phi)(1+\sigma)-\sigma$; $\sigma=0·45$.

The final tensile stress is obtained by adding these three values of T_a weighted in proportion to the areas defined above. The tensile stress has been determined by this method for aluminium up to extensions of about 25 per cent, using the experimental single crystal results of Elam referred to by Taylor (1938). Details of the calculation are listed in Table I., and the resulting polycrystalline stress-elongation curve is shown in fig. 2 together with the experimental points determined by Elam using the same material as that used for the single crystal work. The curve predicted by the Taylor theory is also shown. It will be seen that the curves derived from the Taylor theory and from the present treatment give equally satisfactory agreement with the experimental observations. Although the derivation of a stress-elongation curve must be regarded as an essential part of any deformation theory too much significance should not be attached to the exact agreement of prediction and experiment in view of the various assumptions made. The lack of agreement at higher elongations calls for some comment however. Since the Elam curve is for load-elongation it may be expected to fall below the calculated stress-elongation curves in this region where reduction of cross-sectional area is now no longer negligible. Furthermore, in the calculations random orientation has been assumed throughout, whereas with increasing elongation some slight measure of preferential orientation must occur.

Fig. 2.



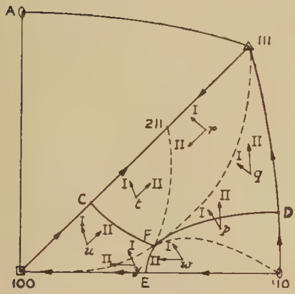
Comparison of predicted tensile stress–elongation curve with Taylor's predicted and experimental curves.

Deformation Texture in Tension.

The development of deformation textures by marked grain rotations implies deformations much more severe than those considered in the stress–strain relationship, where the departure from random orientation due to rotations was not taken into account. It has been proposed above that rotation will occur when the deformation takes place before T_e has reached a corner position. It will be that corresponding to duplex slip if T_e has reached a triangle boundary, and otherwise that corresponding to single crystal slip. In fig. 3 are shown the lines of quickest descent to the corner points, to the [211] direction, and to the maxima on the triangle sides. In addition the rotations for duplex slip are marked with arrowheads along the triangle sides. For an applied stress direction represented by p, the mean path of the effective stress, T_e, is down the contour gradient to reach the [110–111] boundary between [110] and the maximum D. If deformation occurs before T_e reaches this boundary, there is slip on the single most favourable system with consequent rotation along the great circle towards the operative slip direction A. This is shown by the arrow I. If T_e reaches the boundary however, but deformation occurs before it has reached the [110] corner point, rotation corresponding to duplex slip occurs in the direction shown by the arrowhead on the boundary, producing the rotation of p shown by the arrow II. The single and duplex rotations for other typical points q, r, t, u, v and w, are deduced from similar reasoning. As the deformation proceeds, the orientation representing a particular grain (or part of a grain) moves across the unit triangle, and its probable rotations vary according to the particular region in which it is temporarily situated. Thus, with the treatment as oultined so far, it is possible for an orientation such as w to move into the region EF[100], typified by v, then into the region CF[100], and finally into the CF[112] region. One additional feature mentioned in an earlier section will however modify this effect. The grains least likely to rotate are those which require the smallest lateral stresses for T_e to reach the appropriate corner point, *i.e.* those orientations near the corner point. This will apply to the grains near the [111] direction in the DF[111] region, and those near the [100] direction in the EF[100] region. It will not apply to those near [110] in the DF[110] and EF[110] regions or those near [100] in the CF[100] region for the following reason. Consider an orientation similar to p near [110]. Although the effectiveness must travel only a short distance to the [110], any single crystal or duplex rotation which may occur increases the distance from the [110] ,and thus increases the probability of rotation, whereas in the DF[111] region near the [111] any rotation decreases the distance and the probability of rotation. The region between the [211] and the [111] directions is a special case in which, although the rotations lead away from the [111] corner point, the end point of the rotations, the [211], is still sufficiently near this corner point for the probability of rotation to be comparatively low. This should lead to a spread of orientations between [111] and [211], with a tendency for concentration near each of these directions.

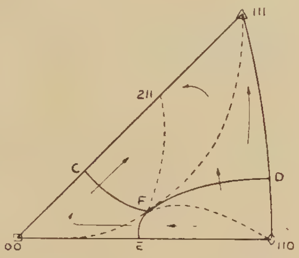
Fig. 4 shows the general trends of the rotations which give rise to the tension texture in fig. 4. There appears to be no experimental work on pure tension textures with which these orientation spreads may be compared, and it has been the custom to consider the textures developed in cold-drawn wires as tension textures. While it is reasonable to suppose that in the wire drawing process longitudinal tension may be the predominating factor, it is probable that the ideal tension texture will be modified by radial compressive stresses, and discussion of this effect will be left until after some consideration of compression textures.

Fig. 3.



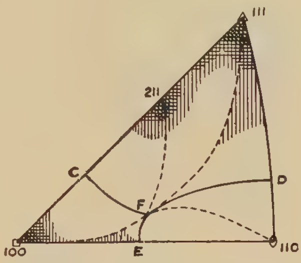
Lines of quickest descent from the shear stress peak (F), and directions of rotation due to single crystal slip (→I) and duplex slip (→II) in a face-centred cubic metal in tension.

Fig. 4.



General trends of the rotations in a face-centred cubic metal in tension.

Fig. 5.



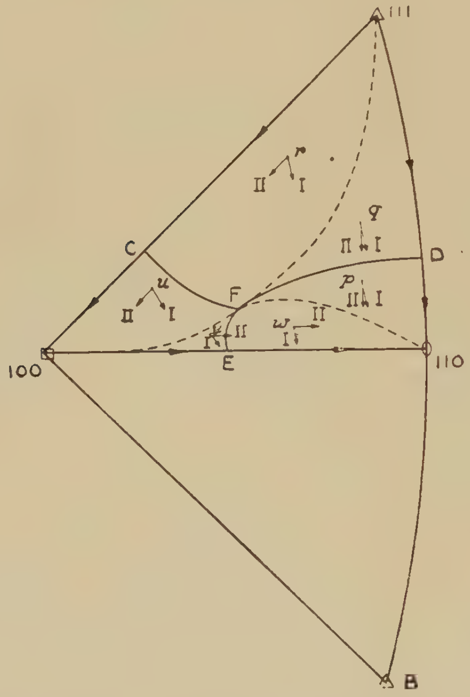
Distribution of grain orientations after tension as represented by their applied stress directions (tension texture).

COMPRESSION TEXTURE.

The compression texture may be derived in an exactly similar manner to that described above, bearing in mind the fact that the rotation directions for single and duplex slip are not simply reversed. For single slip the applied stress direction moves now towards the pole of the operative slip plane, while under duplex slip it moves towards the great circle joining the two operative poles. Thus for an orientation represented by the point *r* in fig. 3 and fig. 6 the single crystal rotation is in the tension case towards the operative [110] slip direction A, and in the compression case towards the pole of the operative (111) slip plane B. The duplex rotations are however merely reversed except in the small region between the [111] and [211] directions.

Considering now the rotations in more detail, it will be seen that the orientations contained in the area F[111][110] will move towards the line F[110] and, in addition, an appreciable proportion of the orientations in CF[111] will also move towards this line. This will clearly denude the regions surounding [111]. Some of those in the CF[111] region, namely, those near to CF, will move across this line. In the region CF[100] there is some tendency for rotation towards the [100] so that a small residual concentration of orientations may be expected near to [100]. Most of the orientations here however, together with those in the EF[100] region, will move in the EF[110] region. In this latter region the path of T_e is primarily along the line E[110], and consequently duplex rotation towards [110] is the more probable. This will also have the effect of turning those orientations which have reached the F[110] boundary from the area above the line sharply towards the [110] direction. Fig. 7 shows these general rotations which would lead to a distribution of orientations shown in fig. 6.

Fig. 6.



Lines of quickest descent from the shear stress peak (F), and directions of rotation due to single crystal slip (→ I) and duplex slip (→ II) in a face-centred cubic metal in compression.

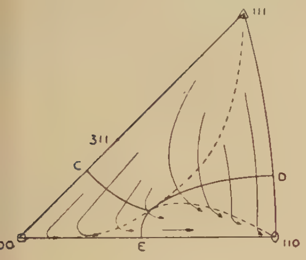
The prediction of a [110] texture is of course amply confirmed by experiments on face-centred metals and the detailed agreement between the observed distribution of orientations, fig. 8 (Barrett and Levenson 1940), and that predicted on the present treatment, is satisfactory.

Rolling Texture.

The simplest combination of stresses for the approximate simulation of the rolling process is that of a compression in the direction of the rolling plane normal together with a tension in the rolling direction (Wever 1931).

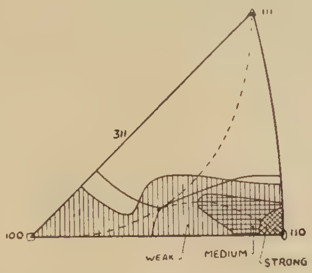
This has been adopted in the present treatment, which examines the ways in which both tension and compression end conditions may be satisfied simultaneously. The resulting rolling texture is therefore not that which would result from a mere super-position of the tension and compression textures.

Fig. 7.



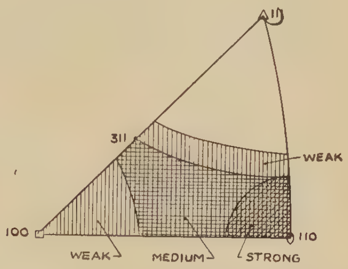
neral trends of the rotations in a face-centred cubic metal in compression.

Fig. 8.



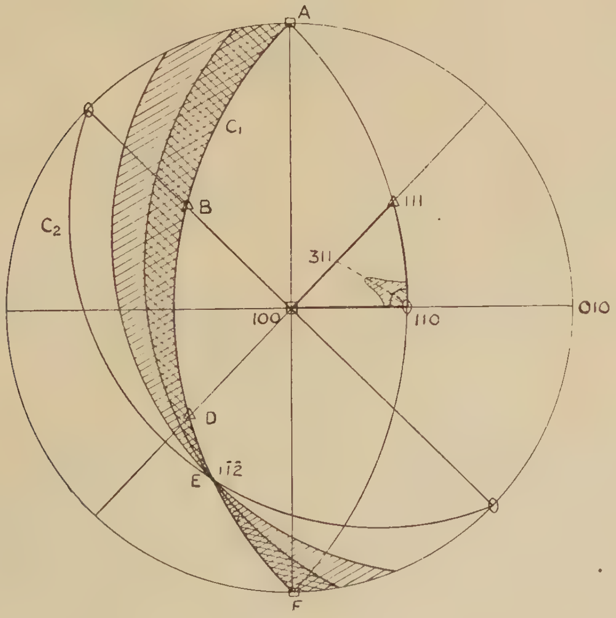
Predicted compression texture for a face-centred cubic metal.

Fig. 9.



Texture in aluminium after 98 per cent compression (Barrett and Levenson 1940).

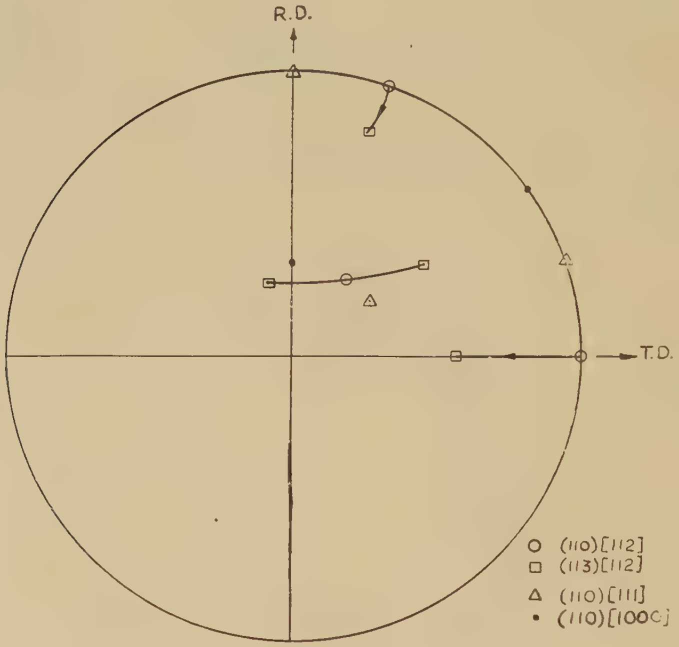
Fig. 10.



Stereographic projection showing the compression texture and the resulting spread in the rolling direction.

The compression normal to the rolling plane will produce a texture such that, for the majority of the crystals, the direction of compression will lie within the shaded area shown in the unit triangle of fig. 10. To simplify the ensuing discussion this texture may be considered as a spread of orientation along the great circle from the [110] direction towards the [311], shown dotted in the figure. Since the rolling direction is perpendicular to the rolling plane normal, the rolling direction must lie on the great circle for which the rolling plane normal is the pole. Thus the rolling direction lies on the circle C_1 for the rolling plane normal in the [110] direction, and on the circle C_2 for the rolling plane normal at the other limit, the [311] direction. The great circles corresponding to the intermediate positions of the compression spread are obtained by oscillation, about the axis [$1\bar{1}\bar{2}$], of a great circle between the limits C_1 and C_2, and the distribution of rolling directions will be as indicated by the cross-hatching between these two great circles. At the same time, the tension in the rolling direction will tend to produce a texture similar to that illustrated in fig. 5, in which most of the crystals have the direction of tension at end-points of the form [111], [211] and [100], the tendency for a spread of orientation about these end positions being ignored at present. Such directions are shown by the points A, B, D, E and F in fig. 10. The only one of these points which completely satisfies both a tension end condition and the compression spread is E, the [$1\bar{1}\bar{2}$] direction. Thus there will be a strong tendency for rotation to occur such that this condition is fulfilled, giving a strong (110)[112] texture. The [111] and [100] end positions of the tension

Fig. 11.

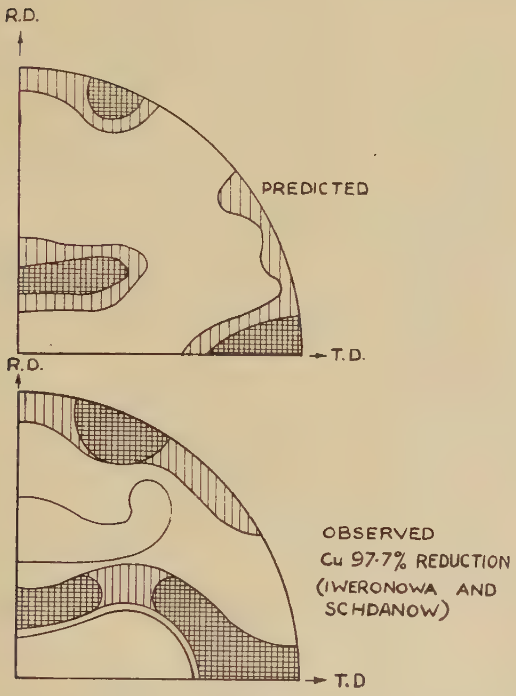


Composite (111) pole figure.

texture satisfy only the [110] end of the compression spread, but since this is the most densely populated region in the compression texture, there will be some tendency towards (110)[111] and (110)[100] components in the final rolling texture, which will be made up of the following preferred orientations.

Direction of tension (rolling direction)	Direction of compression (normal to rolling plane)	Notation
[112]	[110] → [311]	(110) [112] to (311) [112]
[111]	[110]	(110) [111]
[100]	[110]	(110) [100]

Fig. 12.

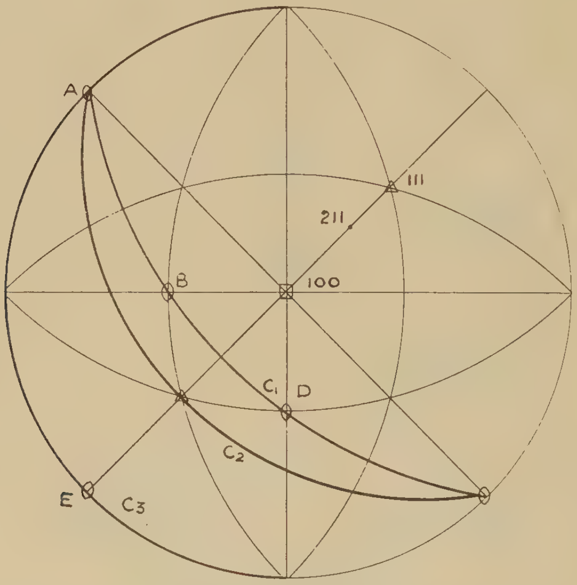


(111) pole figure showing predicted and observed rolling textures.

The ideal (111) pole figure corresponding to these four textures is shown in fig. 11. The lines joining the (111) poles of the (110)[112] and (311)[112] textures are obtained by rotation of the (110)[112] texture about the [112] direction by an amount appropriate to the compression spread from [110] to [311]. In view of the fact that only the poles corresponding to (110)[112] and (311)[112] orientations and the spread between them satisfy both the compression spread and a tension end-position these will predominate. Thus a semi-quantitative pole figure corresponding to these textures with some contributions from the (100)[111] and (110)[100] textures may be obtained as in fig. 12.

On the same figure is shown the (111) pole figure for copper after 97·7 per cent reduction, obtained by Iweronowa and Schdanow (1934). The predicted pole figure shows a much higher degree of preferred orientation, but is otherwise in good agreement with the experimental observations. It is perhaps worth noting that Barrett and Steadman (1942), in an investigation on the rolling of copper using both large grain polycrystalline material and single crystals mounted in polycrystalline blocks, found (*a*) multiple orientations developed in nearly every grain, with the exception of those with initial orientation (110)[112] which were unchanged even after 97·6 per cent reduction, and (*b*) the entire polycrystalline texture developed in a crystal with the initial orientation (100)[001]. This supports the view that the (110)[112] orientation will predominate.

Fig. 13.



Stereographic projection showing ideal tension texture and the corresponding range of compression directions.

DRAWING TEXTURE.

The drawing process will be considered as equivalent to a tension exerted on the metal in the direction of drawing together with uniformly distributed radial stresses in the plane perpendicular to the drawing direction. The longitudinal tension, considered to be the predominating influence, will tend to produce the tension texture already described, namely, concentrations of orientations near [111], [211] and [100] directions. The compression directions, being perpendicular to the tension direction, must be at 90° from these end points, and therefore lie on the great circles C_1, C_2 and C_3 respectively, fig. 13. A compression at any point on these circles attempts to bring a [110] direction, the compression texture being regarded for simplicity as predominantly at [110] texture, into alignment with itself whilst retaining the tension textures. For any point on C_1 this requires a rotation of not more than 30° to the points A, B, and D, and on C_3 of not more than 45° to A and E. On C_2, however, a rotation of as much as 90° to A would apparently be demanded, so it is far more likely that the orientations corresponding to C_2 will, except for those near to A, rotate to join the [110] end points B and D. This will reduce the density of orientations around [211] in the tension texture, and correspondingly augment the [111] texture, thus producing the double fibre texture having [111] and [100] as the fibre axes.

The experimental results of Schmid and Wassermann (1927) show the existence of this double texture in copper, silver and gold, although the proportions of the two textures vary from metal to metal. Aluminium however shows only the [111] component.

It is not possible to predict from purely geometrical considerations these variations in texture, but in the present treatment the development of textures depends on the extent to which the effective stress, T_e, can move before deformation occurs, that is to say, on the ability of the boundary regions of individual grains or sub-units of structure to withstand large stress gradients. This property may well be dependent on the temperature of the experiment in relation to the melting point of the metal and its purity ; the nearer to the melting point and the higher the purity, the less likely is the metal to withstand these stresses. If the movement of T_e is so limited that mainly single crystal rotation occurs the texture will be different from that developed in another face-centred cubic metal in which duplex rotation is the predominating feature. When mainly single crystal rotations occur the [100] component of both the compression and tension textures will be absent. Thus considering the development of the drawing texture, fig. 13, C_3 will be absent and a single fibre axis, the [111], will result. It may be significant that aluminium, which at room temperature is nearer its melting point than the other face-centred metals, copper, silver and gold, does indeed show this single texture.

CONCLUSIONS.

From this qualitative treatment of the deformation of polycrystalline agregates it has been possible to derive for face-centred cubic metals :—

(1) a tensile stress-elongation curve which is in satisfactory agreement with experiment and with that previously derived by Taylor,

(2) tension, compression, rolling, and wire-drawing textures which are in reasonable agreement with observation.

An attempt has been made to give a clearer physical basis of the deformation process and since the above results can be obtained without laborious calculations it may be possible to extend the treatment to body-centred cubic metals.

A very tentative suggestion is put forward to explain the variation of deformation textures among metals of the same crystal structure.

ACKNOWLEDGMENTS.

The authors wish to acknowledge the encouragement and helpful comments received from Dr. N. P. Allen, and the assistance given by the Mathematics Division of the Laboratory in the calculation of the Resolved Shear Stress Tables.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES.

BARRETT, C. S., and LEVENSON, L. H., 1940, *Trans. A.I.M.E.*, **127,** 112.
BARRETT, C. S., and STEADMAN, F. W., 1942, *Trans. A.I.M.E.*, **147,** 57.
BOAS, W., and SCHMID, E., 1931, 1, *Z. tech. Physik*, **12,** 71.
CALNAN, E. A., and BURNS, B. D., 1950, *J. Inst. Metals*, **77,** 445.
GÖLER, V., and SACHS, G., 1929, *Z. Physik*, **55,** 581.
IWERONOWA, W., and SCHDANOW, G., 1934, *Tech. Phys. U.S.S.R.*, **1,** 64.
MISES, R. v., 1928, *Z. angew. Math. Mech.*, **8,** 161.
PICKUS, M. R., and MATHEWSON, C. H., 1939, *J. Inst. Metals*, **64,** 237.
SCHMID, E., and WASSERMANN, G., 1927, *Z. Physik*, **42,** 779.
TAYLOR, G. I., 1938, *J. Inst. Metals*, **62,** 307 ; Stephen Timoshenko, 60th Anniversary Volume, p. 218, Macmillan, New York.
WEVER, F., 1931, *Trans. A.I.M.E.*, **93,** 51.

XCVIII. *The Energy Distribution behind Decaying Shocks.—I. Plane Waves.*

By M. J. Lighthill*.

[Received July 11, 1950.]

Summary.

A hypothesis is made concerning the degree of accuracy of a theory of decaying plane shocks due to Friedrichs. This hypothesis is carefully examined in the light of energy balance ; it is substantiated by the checks made on it, and in the course of the investigation the picture of the disturbance to the fluid as a whole is considerably improved. The work will be used in Part II. to clarify other phenomena involving decaying shocks, including some in steady supersonic flow.

Most of the analysis refers to the following problem. A plane wave, headed by a shock, is produced by a piston, which is moved impulsively into homogeneous stationary fluid and then retarded arbitrarily until it comes to rest. One may usefully ask what is the whole history of the disturbance to the fluid if viscosity and heat conduction be neglected, since the modifying effects of these can be sketched in afterwards. The Friedrichs theory replaces the disturbance by a " simple wave " (with the fluid behind it in its original undisturbed state) and satisfies conditions at the shock neglecting the cube of its strength (see §§2–4 below). A precise hypothesis about the degree of accuracy of this theory is made in §1, but there is no experimental or logical certainty that this is correct at large distances from the piston. A comprehensive check is made as follows, arbitrary thermodynamic properties being assumed for the fluid.

The energy of the simple wave at any time t (§6) is compared with the work done by the piston (§5). The difference (§7) is a quantity of the second order in the initial shock strength, obtained with an error which (on the accuracy hypothesis) is of the third order. This residual energy $R(t)$ can only represent energy of the fluid behind the simple wave (where disturbances are of the third order but their integrated effect may be of the second order). It is comparable with the total entropy gain $S(t)$ of the fluid behind the simple wave, which is known from the theoretical variation of shock strength. In fact the relation

$$R(t)=T\left(1-\frac{1}{2}\frac{T_V P}{TP_V}-\frac{1}{8}\frac{PP_{VV}}{P_V^2}\right)S(t) \quad . \quad . \quad . \quad . \quad (1)$$

is found to hold, neglecting third order terms, where the suffix derivatives are taken with respect to minus the volume at constant entropy (see §2 for notation).

---

* Communicated by the Author.

This rather surprising relation is checked as follows. The residual energy is divided (§ 8) into that due to changes in entropy and pressure separately, the former changes being known. Hence the integrated pressure over the fluid behind the simple wave is determined at all times. Since also these small variations of pressure satisfy the equation of sound (even though the entropy is non-uniform) and the pressure gradient at the stationary piston is zero, this determines completely the third order pressure waves in this part of the fluid. They are such as would be produced if the shock emitted a pressure pulse of magnitude

$$\frac{1}{16A}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)T\frac{dS}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

backwards through the fluid at all times t, which was then reflected at the piston. The bracket in (2) is negative for most fluids but zero for monatomic gases.

The wave (2) is next shown (§ 9) to consist of the reflection of the simple wave at the shock. The reflection coefficient in this process may be calculated exactly by considering the shock conditions with the local equations of motion, but only the third order terms in the reflected wave can be accurate after it has interacted with the simple wave. These are precisely (2), which completes the check by showing that the conclusions of the theory in the large agree with those in the very small.

To demonstrate that the said third-order terms do not change by interaction with the simple wave, the important, but apparently unrecognized, fact is used that, in fairly weak waves, a small variation of entropy is taken account of to a first approximation if the ordinary theory, at constant entropy, is applied with the velocity and *pressure* as independent variables; in particular the flow behind a shock can then be obtained correctly to the third order in the strength. This is explained in § 9.

The extension to an " N-wave ", produced by the piston being retarded through zero to a negative velocity, and then stopping abruptly to produce a second shock, is made in § 10. The accuracy hypothesis, which is again substantiated by the checks made, now implies, however, that the motion of the rear shock after a long time is given less accurately by the theory than that of the front shock. Until the shocks have decayed to a small fraction of their initial strength, the accuracy is unimpaired; also equations (1) and (2) still hold. The pressure waves (2) are now due to the combined effects of reflection of the simple wave at the front shock and transmission of it through the rear shock. But the reflections of these pressure waves at the piston gradually overtake the rear shock as time goes on and changes its path from that given by the Friedrichs theory (equation (79) shows by how much). Thus, when they are rarefaction waves, they communicate their negative energy to the N-wave, by slowing down the rear shock and so widening the N-wave on its low-pressure side. After a very long time, all the residual energy will be due to the increase of total entropy of the fluid with, ultimately, no change in pressure ; *i. e.* $R(\infty)=(1-T_V P/TP_V)S(\infty)$. The remaining energy is propagated to infinity in the N-wave (whose width increases as its amplitude decreases).

It is shown in §11 that, if the piston has stopped on returning to its original position, then after an intermediate length of time the ratios of the energy in the N-wave, in the increased entropy, and in the rear pressure waves, are as 5 : 10 : (—1) for air. The third form of energy is subsequently transferred to the first. On the other hand, if the piston stops much ahead of the original position, the first-named form of energy always predominates.

§1. INTRODUCTION.

DECAYING shocks appear, for example, whenever there is an explosion. Also, when any object moves through fluid faster than sound, the bow shock is observed to become weaker as we follow its course away from and behind the object ; this may also be called decay. A number of papers have been written on the mechanics of the process (see References). Some give detailed theories of its progress in particular cases. These mostly apply only when the shock is fairly weak, and hence only to the later stages of decay for a strong shock. Taylor (1950) treats the early stages of decay for a very strong explosion.

A careful study of the published theories reveals certain doubts, difficulties, and even divergences of opinion. These are connected with considerations of energy and entropy. It is the purpose of this paper, with Part II., to clarify these points, and to answer such questions as (i) When a wave, with a decaying shock in front, and possibly one behind also, passes through fluid, what relationship, if any, is there between the rate of gain of energy, and of entropy, of the fluid behind the wave? (ii) If either is substantial, can a theory of the propagation of the wave, which makes use of assumptions valid only under adiabatic conditions, have any value?

No reference to any but a single type of decaying shock will now be made until §10. The interval will be filled with a very detailed analysis, in this one representative case, of the questions broached above.

Consider a very long tube indeed, of unit cross-sectional area, filled with homogeneous fluid at rest. Let a piston at one end (say $x=0$) begin to move into it impulsively, with a certain finite velocity, at time $t=0$, and be subsequently retarded in an arbitrary manner until after a certain time it is again at rest. Then at any instant, if the tube is long enough, the fluid in motion will be confined between the piston and a plane shock. The shock will decay in strength as it moves along the tube faster than sound.

Viscous stress and heat flow by conduction will be very small, except inside the shock, and where otherwise they are largest (near the front of the wave) will not act on any element of fluid for very long. A theoretical treatment which neglects them, and replaces the very thin shock by a plane surface of discontinuity, should therefore be quite satisfactory. Then the motion is governed by Euler's equations of continuity and momentum, the Rankine-Hugoniot relations at the shock, and constancy of entropy for each element of fluid after the shock has passed it by ; the boundary condition is that the velocity of the fluid in contact with the piston is that of the piston.

No exact mathematical solution is known, even to a particular case of this problem. An approximate theory of Friedrichs * should be good, according to plausible hypotheses, when the velocity of the piston is always rather less than the velocity of sound in the undisturbed fluid, so that the shock is always weak. But there has been no experimental verification, or mathematical proof, of the validity of the approximate theory. It would not perhaps be difficult to produce either if the approximations were required to be good only within a limited distance of the piston. But there are serious difficulties, even in the mathematical discussion, at arbitrarily large distances from the piston, where any terms neglected may have had an arbitrarily long time to influence the propagation of the wave. (It may be noticed here that, the weaker be the shock, the longer does it take to decay to, say, one half of its strength.)

The Friedrichs method is to determine the "simple wave", at the entropy of the undisturbed fluid, which is compatible with the motion of the piston. This is taken as the flow behind the shock. (In a simple wave the velocity and the thermodynamic variables are all functionally related. Any value of one of them is propagated through the fluid at a constant speed, namely at the local speed of sound relative to the local fluid velocity. The simple wave can exist only when the specific entropy is uniform. In the present problem it will terminate with the sound wave emitted by the piston as it comes to rest. Behind this, on simple wave theory, the fluid is once more homogeneous and at rest.) The shock is calculated approximately as a plane surface moving in such a way that the physical quantities in the simple wave thereat take those values which the motion of the shock into fluid at rest would produce, with the cube of the shock strength neglected. The calculations were given by Friedrichs only for the case when the fluid is a perfect gas with constant specific heats, but the extension to a fluid with general thermodynamic properties is a fairly simple matter.

It would be impossible to satisfy the conditions at the shock more exactly ; for example, the specific entropy, which has to be taken uniform in the simple wave, must actually increase at the shock by an amount of order the cube of the strength.

---

* Friedrichs, 1948, pp. 220–231. This theory of the phenomenon, which is the subject of the present §§ 1–9, will be constantly referred to as "the Friedrichs theory", even though in the paper cited many other phenomena were also investigated.

Friedrichs (1948) is cautious about claiming any precise degree of accuracy for this theory, but an attempt at estimating the errors may be made. These errors arise from using the incorrect value of the entropy, and from the fact that only a second order boundary condition can be applied at the shock. These considerations suggest the following

Accuracy Hypothesis.—The errors in the Friedrichs theory are comparable with the changes produced by altering the simple wave in two respects : (i) the specific entropy of the whole wave is altered to its known maximum, *i. e.* by an amount of order the cube of the maximum shock strength ; (ii) the velocity distribution is altered, say by changing the velocity of the piston (continuously, and even after it has actually come to rest) by an amount whose ratio to the velocity of sound is of order the cube of the maximum shock strength.

If this hypothesis is correct, the Friedrichs theory is of considerable value at all distances from the piston, as will be seen in §4. Even after many multiples of the duration of motion of the piston, the position of the shock (relative to its position if it has been an acoustic wave) is in error only by a factor one plus a term of order the square of the maximum shock strength. Even though this factor multiplies a quantity which grows indefinitely, the approximation may be regarded as highly satisfactory.

The accuracy hypothesis might be questioned on two grounds, both based on the fact that the total entropy gain of the fluid as a whole, as a result of the passage of the shock, is of order the square of the initial shock strength *. (i) The effect on the wave of the terms neglected, if it accumulates as time goes on, might be held as a consequence to be of order greater than the cube of the maxiumm shock strength. (ii) It might be expected that the total energy of the fluid is defective (on the Friedrichs theory) by an amount of order the square of the maximum strength (and this is true), and that this indicates errors of such an order in the wave.

The author rejects objection (i) on the grounds that the form of the accuracy hypothesis already permits propagation of errors along the diverging characteristics, and hence their progressive accentuation ; thus, in the model simple wave used for comparison, the effects of the new value for the specific entropy accumulate as time goes on. The defective total energy mentioned in (ii) is the subject of this paper, which will analyse it in detail, assuming the accuracy hypothesis. It is sufficient here to remark that, though errors within the simple wave itself could not account for this deficiency in total energy, the region behind it (which the Friedrichs theory gives as undisturbed) is permitted, on the accuracy hypothesis, to include disturbances of order the cube of the maximum shock strength. The integrated effect of these over the extensive region behind the simple wave is sufficient to account for the discrepancy in energy.

---

* That this is so, although the gain in specific entropy of any element is of order the cube of the strength, is because the weaker shocks take longer to decay.

But the hypothesis remains unproved, and it was to make a further test of its correctness, and in particular to make a detailed critique of objection (ii), that the further deductions from the hypothesis, given below, were made. If they had led to a contradiction, or to physically unplausible results, the hypothesis would have had to be modified. In fact, the results are neither impossible or unlikely, and the considerably fuller picture of the flow as a whole, which this investigation in the large provides, is confirmed in quantitative detail by an investigation of the local reflection of the simple wave at the shock.

The deductions which are to be made have already been sketched in the Summary. At any time t the total increase in energy, kinetic and internal, of the fluid in the simple wave over its internal energy in the undisturbed state—briefly, " the energy in the simple wave "—will be calculated on the accuracy hypothesis. Now the energy in the whole fluid must equal the work done by the piston. Hence, if the accuracy hypothesis is true, the difference must represent energy of the fluid behind the simple wave. From the expression for this residual energy, the pressure distribution in this region is deduced, to a first approximation, as a third order function varying only relatively gradually. The same pressure distribution is then deduced in a totally different manner, as a result of reflection of the simple wave at the shock followed by reflection at the piston.

The author has thought it worth while to set out the work for a fluid with perfectly general thermodynamic properties, for three reasons:—(i) The proportionality of residual energy and entropy shown in equation (1) (in the absence of which the theory might find itself in difficulties), and the agreement in expression (2) as obtained by two methods, then appear as the result of the cancellation of a large number of terms, in a manner impossible to believe fortuitous. (ii) The work shows up explicitly what terms arising in it are due to departures from adiabatic conditions * (roughly those bringing in the temperature). (iii) It is desirable to place on record a statement of the Friedrichs theory for fluids other than perfect gases.

The " tail " behind the simple wave is found to be a compression wave, in which the slightly reduced pressure directly behind the simple wave is gradually made up to its undisturbed value. Technically this compression wave must ultimately form a shock, but it is so weak that this could not happen for a very long time, and need hardly be considered in detail †. One may also notice that the increased entropy in the general neighbourhood of the piston, which the theory leaves as permanent, would gradually be flattened out, for a real fluid, by conduction of heat, causing a further slight increase in the total entropy, though not in the total energy.

---

* See § 3 for a discussion of this point, and of the errors in a paper of Martin due to his ignoring it.

† The indication that an N-wave must ultimately form, even in the present problem, is, however, interesting.

§2. THE SIMPLE WAVE.

The notation to be used for the various physical quantities, at points in the undisturbed fluid and elsewhere, is indicated in the table below.

Quantity	Undisturbed	Disturbed
Pressure	P	$P+p$
Specific volume	V	$V-v$
Velocity	0	u
Velocity of sound	$A=V\sqrt{P_V}$	$a=(V-v)\sqrt{(\partial p/\partial v)_s}$
Temperature	T	.....
Internal energy	E	$E+e$
Specific entropy	0	s

Derivatives, in states not that of the undisturbed fluid, will be indicated in the manner usual in thermodynamics ; thus $(\partial p/\partial v)_s$ indicates a derivative at constant entropy. But the values in the undisturbed state of derivatives along an adiabatic (*i. e.* at constant entropy) will be denoted by suffixes and by the use of capital letters ; P_V is the value of the derivative just mentioned in the undisturbed fluid, P_{VV} the corresponding second derivative, and so on *. Suffix S will also be used to denote differentiation with respect to entropy at constant volume in the undisturbed state.

The Euler equations of continuity and momentum are

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+(V-v)\frac{\partial u}{\partial x}=0, \qquad \frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+(V-v)\frac{\partial p}{\partial x}=0. \quad . \quad (3)$$

Hence at constant entropy

$$\left.\begin{aligned}\frac{a}{V-v}\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}\right)-\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}\right)+a\left(\frac{\partial u}{\partial x}-\frac{a}{V-v}\frac{\partial v}{\partial x}\right)=0,\\ \frac{a}{V-v}\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}\right)+\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}\right)+a\left(\frac{\partial u}{\partial x}+\frac{a}{V-v}\frac{\partial v}{\partial x}\right)=0.\end{aligned}\right\} \quad . \quad (4)$$

A solution is therefore given by putting

$$u=\int_0^v a\frac{dv}{V-v} \quad . \ . \ . \ . \ . \ . \ . \ . \ . \quad (5)$$

to satisfy the first equation, so that the second becomes

$$\frac{\partial u}{\partial t}+(u+a)\frac{\partial u}{\partial x}=0, \quad . \ . \ . \ . \ . \ . \ . \ . \ . \quad (6)$$

showing that u is constant along straight characteristics $dx/dt=u+a$. This solution is the " simple wave ".

---

* Notice that P_V is therefore *minus* the derivative of pressure with respect to volume (and P_{VVV} minus the third derivative, but P_{VV} plus the second derivative). This is convenient as making the quantities all positive for most fluids.

Friedrichs showed that it was convenient to regard the simple wave as specified by a single function $\sigma(\xi)$, which he defined as follows. ξ is a parameter constant on each straight characteristic, namely the value of x where the characteristic produced would meet the axis $t=0$. It is always negative in the problem of this paper. Then, writing $u+a=\mathrm{A}(1+\sigma)$, it follows, since u and a are constants on any characteristic, that σ and ξ are functionally related. Actually, in the present problem, it is more convenient to regard ξ as a function of σ than vice versa, for it is a problem in which this function is one-valued, unlike some considered by Friedrichs. Then the equation of a characteristic is

$$x=\xi(\sigma)+\mathrm{A}t(1+\sigma). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Since σ is a non-dimensional measure of the disturbance to the fluid, its cube may always be neglected in the Friedrichs theory.

All the physical quantities are related in the simple wave ; all will now be expressed in terms of σ, neglecting cubes of disturbances. Firstly, by (5),

$$u=\int_0^v \frac{\mathrm{A}+\mathrm{A_V}v}{\mathrm{V}-v}\,dv=\frac{\mathrm{A}}{\mathrm{V}}v+\frac{1}{2}\left(\frac{\mathrm{A}}{\mathrm{V}^2}+\frac{\mathrm{A_V}}{\mathrm{V}}\right)v^2, \quad . \quad . \quad . \quad . \quad (8)$$

whence

$$\sigma=\frac{u+a-\mathrm{A}}{\mathrm{A}}=\left(\frac{1}{\mathrm{V}}+\frac{\mathrm{A_V}}{\mathrm{A}}\right)v+\frac{1}{2}\left(\frac{1}{\mathrm{V}^2}+\frac{\mathrm{A_V}}{\mathrm{AV}}+\frac{\mathrm{A_{VV}}}{\mathrm{A}}\right)v^2. \quad . \quad . \quad . \quad (9)$$

Replacing A by $\mathrm{VP}^{\frac{1}{2}}_{\mathrm{V}}$, and its derivatives accordingly, (9) may be rewritten

$$\sigma=\frac{\mathrm{P_{VV}}}{2\mathrm{P_V}}v-\frac{1}{2}\left(\frac{\mathrm{P_{VV}}}{2\mathrm{VP_V}}-\frac{\mathrm{P_{VVV}}}{2\mathrm{P_V}}+\frac{\mathrm{P^2_{VV}}}{4\mathrm{P^2_V}}\right)v^2, \quad . \quad . \quad . \quad . \quad . \quad (10)$$

whence
$$v=\frac{2\mathrm{P_V}}{\mathrm{P_{VV}}}\left[\sigma+\left(\frac{1}{2}+\frac{\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}\right)\sigma^2\right] \quad . \quad . \quad . \quad . \quad (11)$$

and
$$u=\frac{2\mathrm{AP_V}}{\mathrm{VP_{VV}}}\left[\sigma+\left(1+\frac{\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}\right)\sigma^2\right]. \quad . \quad . \quad . \quad . \quad (12)$$

The coefficient of σ^2 in u vanishes in the case considered by Friedrichs, of a perfect gas with constant specific heats.

According to the accuracy hypothesis the errors in (7), (11) and (12) are such as would be produced by adding terms of order σ^3 (or rather the maximum of this quantity) to those in σ and σ^2.

§ 3. Conditions at the Shock.

The values of disturbed quantities behind the shock are related by equations expressing conservation of mass momentum and energy for an element of fluid as the shock passes through it, say with velocity U. These are

$$\frac{\mathrm{U}}{\mathrm{V}}=\frac{\mathrm{U}-u}{\mathrm{V}-v}, \qquad p=\frac{\mathrm{U}}{\mathrm{V}}u, \qquad (\mathrm{P}+p)u=\frac{\mathrm{U}}{\mathrm{V}}\left(\tfrac{1}{2}u^2+e\right). \quad . \quad . \quad (13)$$

Elimination of U and u gives the Hugoniot relation between the thermodynamic variables

$$(P+\tfrac{1}{2}p)v=e. \qquad \dots\dots\dots\dots \qquad (14)$$

It is convenient to expand (14) in series, using that $P+p=(\partial e/\partial v)_s$. The best form is as

$$[E_V+\tfrac{1}{8}(E_{VVV}v^2+2E_{VVS}vs+E_{VSS}s^2)]v=E_Vv+E_Ss$$
$$+\frac{1}{24}(E_{VVV}v^3+3E_{VVS}v^2s+3E_{VSS}vs^2+E_{SSS}s^3)+O(v^5+s^5), \quad . \quad (15)$$

where (in this one equation only in the whole paper) suffix derivatives refer not to the undisturbed state but to a state with specific volume and entropy halfway between those ahead and of behind the shock. (In this way quadratic and quartic terms on both sides of the equation are avoided.) This gives the entropy as

$$s=\frac{E_{VVV}}{12E_S}v^3=\frac{P_{VV}}{12T}v^3, \qquad \dots\dots\dots \qquad (16)$$

the error being $O(v^5)$ if P_{VV} and T are taken in the said intermediate state, but $O(v^4)$ if (as will be sufficient in what follows) they are taken in the undisturbed state.

Equation (16) is a convenient form of the relation between the thermodynamic variables behind the shock: it states how near to the adiabatic this relation is. The velocities u and U are deducible, by (13), as

$$u=(pv)^{\frac{1}{2}}, \qquad U=V(p/v)^{\frac{1}{2}}. \qquad \dots\dots\dots \qquad (17)$$

It should be noticed that it follows from these equations, with (16), that in the expansion of u in powers of p or v alone, neglecting cubes of the disturbances, the adiabatic relation may be used, but that this is not so in the expansion of U, which is

$$U=V[P_V+\tfrac{1}{2}P_{VV}v+\tfrac{1}{6}P_{VVV}v^2+\dots+P_S(s/v)+\dots]^{\frac{1}{2}}$$
$$=A\left[1+\frac{P_{VV}}{2P_V}v+\left(\frac{P_{VVV}}{6P_V}+\frac{T_V}{P_V}\frac{P_{VV}}{12T}\right)v^2\right]^{\frac{1}{2}}, \qquad \dots\dots \qquad (18)$$

since $P_S=E_{VS}=T_V$.

Martin (1947) falls into the error of neglecting the effect on U of the entropy change. Hence second order terms in his calculations are wrong. One result of this mistake is that he obtains a reflection of a simple wave at a shock, whereas in fact, for the special case of a perfect gas with adiabatic index 5/3 which he considers, the reflection coefficient is at most of the fourth order, as will be seen in § 9. Friedrichs avoids the mistake by using the explicit solution for a perfect gas of the Rankine-Hugoniot relation. The T_V/T in (18), which will recur in many subsequent formulae, is a constant reminder that in one place, at least, the Friedrichs theory does have to take account of the entropy change at the shock.

§ 4. DETERMINATION OF THE SHOCK.

The shock condition must now be expressed, using the "simple wave" values (§2) for the various physical quantities behind it. Then v is related to the fundamental variable σ by (11), and the expansion (18) becomes

$$\mathrm{U}=\mathrm{A}(1+\tfrac{1}{2}\sigma+\lambda\sigma^2),$$

where
$$\lambda=\frac{1}{8}+\frac{\mathrm{P_V}}{6\mathrm{P_{VV}}}\left(\frac{3}{\mathrm{V}}-\frac{\mathrm{P_{VVV}}}{\mathrm{P_{VV}}}+\frac{\mathrm{T_V}}{\mathrm{T}}\right). \qquad (19)$$

For a perfect gas with constant specific heats λ is simply $\frac{1}{8}$.

At time t let the variable σ take the value σ_1 at the shock, so that σ_1 is a measure of the strength of the shock at time t. Then the condition $dx/dt=\mathrm{U}$ along the shock, with (7) and (19), gives

$$(\xi'(\sigma_1)+\mathrm{A}t)\frac{d\sigma_1}{dt}+\mathrm{A}(1+\sigma_1)=\mathrm{A}(1+\tfrac{1}{2}\sigma_1+\lambda\sigma_1^2), \qquad (20)$$

whence
$$(\tfrac{1}{2}\sigma_1-\lambda\sigma_1^2)\mathrm{A}\frac{dt}{d\sigma_1}+\mathrm{A}t=-\xi'(\sigma_1). \qquad (21)$$

The initial condition for the shock is $x=0$ at time $t=0$, and so by (7) $\xi=0$ when $t=0$. If $\sigma=\sigma_0$ initially, so that σ_0 is a measure of the maximum shock strength, then (21) may be solved exactly as

$$\mathrm{A}t=2\left(\frac{1-2\lambda\sigma_1}{\sigma_1}\right)^2\int_{\sigma_1}^{\sigma_0}\frac{\sigma\,d\xi(\sigma)}{(1-2\lambda\sigma)^3}, \qquad (22)$$

but, since cubes of σ have been neglected in the formulæ on which this equation is based (without which it would be impossible to satisfy the shock conditions completely) the only significant terms in (22) are

$$\mathrm{A}t=\left(\frac{2}{\sigma_1^2}-\frac{8\lambda}{\sigma_1}\right)\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi(\sigma)+\frac{12\lambda}{\sigma_1^2}\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi(\sigma). \qquad (23)$$

This tells us the time t at which the non-dimensional measure of disturbance σ takes the value σ_1 just behind the shock. It is easy to verify that the condition (17) on u is also satisfied to the required order, by comparing it with (5).

Here the significance of the errors in (23) which the accuracy hypothesis permits should be noticed. They are terms like

$$\sigma_1^{-2}\int_{\sigma_1}^{\sigma_0}\sigma^3\,d\xi, \qquad \sigma_1^{-1}\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi, \qquad \int_{\sigma_1}^{\sigma_0}\sigma\,d\xi, \text{ etc.}$$

Their ratio to the terms included is always of order σ_0^2, although the first two become indefinitely large as time goes on and σ_1 decreases towards zero. In consequence, the value of σ_1 for given t is also in error only by a factor $1+\mathrm{O}(\sigma_0^2)$, as stated in §1.

§ 5. THE PISTON.

It is necessary to have an expression for the motion of the piston in terms of the fundamental function $\xi(\sigma)$. The former should logically come first, as being the cause of the whole motion, but it is much more convenient to work from σ throughout. It will be sufficient for our purposes to neglect the square of σ in determining the time t at which the piston emits the characteristic specified by σ.

At the piston $dx=u\,dt$; hence, by (7) and (12),

$$d\sigma(\xi'(\sigma)+\mathrm{A}t)+\mathrm{A}\,dt\left(1+\sigma\left(1-\frac{2\mathrm{P_V}}{\mathrm{VP_{VV}}}\right)\right)=0, \quad . \quad . \quad . \quad (24)$$

whose solution, still neglecting σ^2, is

$$\mathrm{A}t=-\xi(\sigma)+\sigma\xi(\sigma)+\frac{2\mathrm{P_V}}{\mathrm{VP_{VV}}}\int_{\sigma}^{\sigma_0}\sigma\,d\xi, \quad . \quad . \quad . \quad . \quad (25)$$

taking $x=0$, and so $\xi=0$, initially (when $\sigma=\sigma_0$).

Where the piston comes to rest $u=0$, and hence, by (12), $\sigma=0$ (neglecting, on the accuracy hypothesis, σ_0^3). The "simple wave" will be taken as the region satisfying $0<\sigma<\sigma_1$ exactly, *i. e.* lying between the shock and the characteristic whose speed is that of sound in the undisturbed fluid. This characteristic is not emitted precisely as the piston comes to rest, but (as will be seen in § 9) very slightly earlier.

The work done by the piston against the pressure of the fluid, neglecting σ_0^3, is

$$\int_{\sigma_0}^{0}(\mathrm{P}+p)u(dt/d\sigma)\,d\sigma$$

$$=\int_{0}^{\sigma_0}(\mathrm{P}+\mathrm{P_V}v)\frac{2\mathrm{AP_V}}{\mathrm{VP_{VV}}}\left(\sigma+\left(1+\frac{\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P_{VV}^2}}\right)\sigma^2\right)\frac{1}{\mathrm{A}}$$

$$\times\left(\xi'(\sigma)-\xi(\sigma)-\left(1-\frac{2\mathrm{P_V}}{\mathrm{VP_{VV}}}\right)\sigma\xi'(\sigma)\right)d\sigma$$

$$=\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left[\int_{0}^{\sigma_0}\sigma\,d\xi+\left(\frac{1}{2}+\frac{2\mathrm{P_V^2}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P_{VV}^2}}\right)\int_{0}^{\sigma_0}\sigma^2\,d\xi\right], \quad . \quad . \quad . \quad . \quad (26)$$

where (12) and (25) have been used and, in integrating $\sigma\xi(\sigma)$ by parts, the fact that it vanishes at both limits.

To the order of approximation employed, (26) must represent the total gain in energy of the fluid at any time after the piston has come to rest.

§ 6. ENERGY OF THE SIMPLE WAVE AT TIME t.

The energy of the simple wave will now be calculated, on the accuracy hypothesis, at any time t after the piston has come to rest. Under adiabatic conditions the increased kinetic and internal energy per unit mass of a disturbed element of fluid is

$$\tfrac{1}{2}u^2+\int_{0}^{v}(\mathrm{P}+p)\,dv,$$

which in the simple wave is

$$\tfrac{1}{2}u^2+Pv+\tfrac{1}{2}P_V v^2=\frac{2PP_V}{P_{VV}}\left[\sigma+\left(\frac{1}{2}+\frac{P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}+\frac{2P_V^2}{PP_{VV}}\right)\sigma^2\right], \qquad (27$$

neglecting σ^3.

The total energy in the simple wave $0<\sigma<\sigma_1$, obtained by integrating (27) multiplied by $dx/(V-v)$, with x given by (7), is

$$\frac{2PP_V}{P_{VV}}\int_0^{\sigma_1}\left[\sigma+\left(\frac{1}{2}+\frac{P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}+\frac{2P_V^2}{PP_{VV}}\right)\sigma^2\right]\left(\frac{1}{V}+\frac{2P_V}{V^2P_{VV}}\sigma\right)$$
$$\times(\xi'(\sigma)+At)\,d\sigma=\frac{2PP_V}{VP_{VV}}\left[\int_0^{\sigma_1}\sigma\,d\xi+At\tfrac{1}{2}\sigma_1^2+\left(\frac{1}{2}+\frac{2P_V^2}{PP_{VV}}\right.\right.$$
$$\left.\left.+\frac{3P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}\right)\left(\int_0^{\sigma_1}\sigma^2\,d\xi+At\tfrac{1}{3}\sigma_1^3\right)\right]. \quad . \quad (28)$$

With At expressed in terms of σ_1 by (23), this energy becomes

$$\frac{2PP_V}{VP_{VV}}\left[\int_0^{\sigma_0}\sigma\,d\xi-4\lambda\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi+6\lambda\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi\right.$$
$$\left.+\left(\frac{1}{2}+\frac{2P_V^2}{PP_{VV}}+\frac{3P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}\right)\left(\int_0^{\sigma_1}\sigma^2\,d\xi+\tfrac{2}{3}\sigma_1\int_0^{\sigma_1}\sigma\,d\xi\right)\right] \quad . \quad (29)$$

Since in (28) At is multiplied entirely by terms σ_1^2, σ_1^3, etc. *, the error in applying (23) to obtain (29) is at most of order σ_0^3 (arising from integrals like $\int_0^{\sigma_0}\sigma^3\,d\xi$). Thus, on the accuracy hypothesis, (29) is the correct first and second order expression for the energy of the simple wave.

§7. THE RESIDUAL ENERGY.

The difference between (26) and (29) must therefore represent energy in the region behind the simple wave. Subtracting, one obtains for this residual energy

$$R(t)=\frac{2PP_V}{VP_{VV}}\left(\frac{1}{2}+\frac{2P_V^2}{PP_{VV}}+\frac{3P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}-6\lambda\right)$$
$$\times\left(\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi-\tfrac{2}{3}\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi\right). \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Equation (30) has been proved only at times t after the piston has come to rest. But, before this, $At=O(1)$, and study of (23) shows that then $|\xi(\sigma_1)|=O(\sigma_0)$, so that (30) is $O(\sigma_0^3)$. The residual energy at this early stage may also be shown to be of the third order, and so (30) remains true in the sense that the difference between its left- and right-hand sides is always a third order quantity.

---

* This is because the simple wave has been considered as terminating with the characteristic on which $\sigma=0$ exactly. This point will be examined further in § 9.

It will be observed that all integrals in (26) and (29) involving the range $0<\sigma<\sigma_1$ have disappeared by the cancellation of eight distinct terms. This suggests that $R(t)$ depends only on the properties of the *shock* previous to time t. The total entropy of the fluid at time t, say $S(t)$, also has this property, and a comparison of the two is clearly desirable.

Unit mass of fluid, on passing through the shock, gains an entropy s given by (16), and subsequently retains it. To a first approximation mass passes through the shock at a rate A/V; also v is given by (11) as $2P_V\sigma/P_{VV}$. Hence the rate of increase of total entropy of the fluid is

$$\frac{dS}{dt}=\frac{A}{V}\frac{P_{VV}}{12T}\left(\frac{2P_V}{P_{VV}}\right)^3\sigma_1^3. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

Hence, using the first approximation to (23) to integrate (31),

$$S(t)=\int_{\sigma_0}^{\sigma_1}\frac{dS}{dt}\frac{dt}{d\sigma_1}d\sigma_1=\frac{A}{V}\frac{P_{VV}}{12T}\left(\frac{2P_V}{P_{VV}}\right)^3\int_{\sigma_0}^{\sigma_1}\sigma_1^3\frac{1}{A}\left(-\frac{2\xi'(\sigma_1)}{\sigma_1}-\frac{4}{\sigma_1^3}\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi\right)d\sigma_1$$

$$=\frac{2}{3}\frac{P_V^3}{VTP_{VV}^2}\left(6\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi-4\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi\right). \qquad . \quad . \quad . \quad (32)$$

Comparison of (32) with (30) shows that $R(t)$ and $S(t)$ are proportional at all times (neglecting third order terms). Substitution for λ in (30) from (19), and division, gives their relationship in the form already stated in the summary, namely

$$R(t)=T\left(1-\frac{1}{2}\frac{T_VP}{TP_V}-\frac{1}{8}\frac{PP_{VV}}{P_V^2}\right)S(t). \qquad . \quad . \quad . \quad (1\text{ bis})$$

The " coincidences " necessary to obtain this relatively simple relationship already add to one's confidence in the Friedrichs theory. On the other hand, the relationship is not so simple that any rough-and-ready arguments, based on familiarity with the thermodynamics of systems in equilibrium, would have been able to predict it.

§8. STATE OF THE FLUID BEHIND THE SIMPLE WAVE.

According to the accuracy hypothesis, the velocities, and the disturbances to the thermodynamic variables, are of the third order in the region behind the simple wave. The total energy is $R(t)$. Since kinetic energy terms are of the sixth order, they are negligible in building up the total energy, and so

$$R(t)=\frac{1}{V}\int_0^{At}e\,dx. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Here the limits are not exactly correct for describing the region behind the simple wave, but the error involved in using them in (33) is still of the third order, since the true boundaries are exactly parallel to $x=0$ and $x=At$ respectively.

Now, to a first approximation,

$$e=\mathrm{T}s+\mathrm{P}v=\mathrm{T}s+\mathrm{P}\left(-\frac{\mathrm{P_S}}{\mathrm{P_V}}s+\frac{1}{\mathrm{P_V}}p\right). \quad . \quad . \quad . \quad . \quad (34)$$

Hence

$$\int_0^{\mathrm{A}t} p\,dx=\frac{\mathrm{P_V}}{\mathrm{P}}\left\{\mathrm{VR}(t)-\left(\mathrm{T}-\frac{\mathrm{PT_V}}{\mathrm{P_V}}\right)\mathrm{VS}(t)\right\}$$

$$=\tfrac{1}{8}\mathrm{T}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\mathrm{S}(t) \quad . \quad . \quad . \quad . \quad . \quad (35)$$

by (34), (33) and (1). Thus the fluid behind the simple wave cannot be entirely at its original undisturbed pressure unless the bracket in (35) is zero (so that the bracket in (1) becomes $1-\mathrm{T_V P/TP_V}$) *.

For a perfect gas with constant adiabatic index γ the bracket in (35) is $3\gamma-5$. Thus it vanishes for monatomic gases, but is negative for other gases, even if they are only approximately perfect, since for these γ is always considerably less than $\frac{5}{3}$. The bracket is probably not positive for any fluid. Thus for liquids the second term, corresponding to the relative decrease of compressibility with decreasing specific volume, is always considerable, exceeding about 3 even at high pressures, and at least 6 at normal pressures. On the other hand, the first term may be written, in terms of the coefficient of expansion α and the compressibility κ, as $4\alpha\mathrm{V}/c_v\kappa$. Here $4\mathrm{V}/c_v$ is limited by the law of Dulong and Petit to less than about 3×10^{-7} cm.^{2}/dyne.deg. But α/κ, the pressure increase due to raising a rigidly enclosed gram of liquid through one degree, is not as much as twenty atmospheres at normal pressures, as (it is seen) would be needed to make the first term in the bracket attain to the lower limit 6 of the second term. On the other hand, at high pressures the coefficient of expansion falls off rapidly, while the compressibility increases, indicating that here too the bracket in (35) is negative. It will be assumed, throughout what follows, that the bracket is non-positive †, as has here been indicated may be true for all, or nearly all, fluids. Then, by (35), the average pressure behind the simple wave is below the undisturbed value (or equal to it, to the third order, for a monatomic gas, a gloss which will not be repeated henceforth in similar contexts).

Now in this region of small variations of pressure, entropy and velocity, those of *pressure* at least are governed by the equation of sound. For by (3), with small disturbances,

$$\frac{\partial v}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}=0, \qquad \frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial p}{\partial x}=0. \quad . \quad . \quad . \quad . \quad . \quad (36)$$

---

* Nor can the fluid be at its original undisturbed specific volume, which would mean that (1) is replaced by $\mathrm{R}(t)=\mathrm{TS}(t)$.

† In words, the conjecture is that the slope of an adiabatic, as graph of pressure against volume, increases at least as rapidly as the fourth power of the temperature.

But entropy is effectively independent of t, because constant for a given fluid element. Hence $\partial v/\partial t = (V^2/A^2)\partial p/\partial t$, and, therefore,

$$\frac{\partial^2 p}{\partial t^2} = A^2 \frac{\partial^2 p}{\partial x^2}. \tag{37}$$

Also, by the second of equations (3), at the piston (where $x \simeq 0$, and $u=0$ at all times t after the piston has come to rest),

$$(\partial p/\partial x)_{x=0} = 0. \tag{38}$$

Equations (35), (37) and (38) determine p completely (to the present third order approximation). By (37) and (38)

$$p = f(At+x) + f(At-x) \tag{39}$$

for some f. Equation (35) then gives

$$\int_0^{2At} f(y)\,dy = \tfrac{1}{8}T\left(4\frac{VT_V}{T} - \frac{VP_{VV}}{P_V}\right)S(t). \tag{40}$$

Hence, behind the simple wave,

$$p = \frac{T}{16A}\left(4\frac{VT_V}{T} - \frac{VP_{VV}}{P_V}\right)\left[S'\left(\frac{1}{2}\left(t+\frac{x}{A}\right)\right) + S'\left(\frac{1}{2}\left(t-\frac{x}{A}\right)\right)\right], \tag{41}$$

where $S(t)$ is given by (31) *.

The first term in square brackets in (41) represents a pressure wave propagated in the direction x decreasing, opposite to that of the shock. The second term is the reflection of the first at the piston. At $x=At$ the first is simply $S'(t)$. Since the shock satisfies $x=At$ with a proportional error which is uniformly small, we may say that the pressure fluctuations behind the simple wave are such as would be produced if at each time t the shock sent backwards a pressure pulse

$$\frac{T}{16A}\left(4\frac{VT_V}{T} - \frac{VP_{VV}}{P_V}\right)\frac{dS}{dt} = \frac{P_V^3}{24VP_{VV}^2}\left(4\frac{VT_V}{T} - \frac{VP_{VV}}{P_V}\right)\sigma_1^3, \tag{42}$$

as stated in the Summary, which is subsequently reflected at the piston.

Since $S'(t)$ decreases monotonically to zero, it follows that after a long time the pressure (41) is only considerable within a fringe or "tail", say $A(t-t_1) < x < At$, behind the simple wave, wherein $S'(\frac{1}{2}(t-x/A))$ is still significant. Thus only the reflected wave is of importance after a long time. The main simple wave "over-expands" slightly to a negative pressure

$$\frac{T}{16A}\left(4\frac{VT_V}{T} - \frac{VP_{VV}}{P_V}\right)S'(0), \tag{43}$$

---

* It may be remarked that, by (36), the velocity takes a form similar to (41), but with an additional factor V/A and with the sign of the first term in square brackets changed.

for large t (and to twice this value for small t), and this pressure is recovered during the passage of a very much wider compression wave. At every point

$$\lim_{t\to 0} p=0 ;$$

thus the initial undisturbed pressure is finally recovered. The entropy, however, remains permanently altered, unless, as pointed out at the end of § 1, heat conduction is taken into account.

§ 9. CAUSE OF THE BACKWARD-EMITTED PRESSURE WAVE.

It has been inferred, by a somewhat roundabout investigation of the disturbance to the fluid as a whole, that the shock emits a pressure wave backwards, the pulse emitted at time t being given by (2), or (42). This will now be verified directly by a study of conditions just behind the shock. This will show that the pressure wave may be regarded as a reflection of the simple wave at the shock.

Immediately behind a plane shock, which moves into undisturbed fluid, the pressure may locally be divided into two waves each moving with the local speed of sound relative to the local fluid velocity, one in the direction of motion of the shock and the other in the opposite direction. The ratio of the pressure gradient across the latter to that across the former may be called the reflection coefficient of waves at the shock. It may be calculated exactly for shocks of arbitrary strength, and is proportional to the cube of the strength for weak shocks. It does not seem to have been given explicitly in the literature, though an essentially similar problem, of two-dimensional steady supersonic flow, was solved by the author (Lighthill 1949 a).

The reflected pressure wave is of the same order of magnitude as the entropy variations behind the shock. However, it remains unchanged to this order as it passes through the simple wave into the fluid behind. This is because *pressure and velocity*, unlike the other variables involved, *remain insensitive to variations of entropy to a first approximation*—i.e. *ignoring products of variations of entropy with those of other quantities.*

For sound waves this fact is known, and has been used in § 8 ; but this section is concerned with waves of finite, though not large, amplitude. Special cases of the result have been used by the author in his work on the flow behind a stationary shock (Lighthill 1949 a) and on the diffraction of blast (Lighthill 1949 b, 1950), but he has realized its truth in the generality given above only more recently.

The proof fills only one paragraph. In the full equations of continuity and momentum for an inviscid fluid, derivatives of the pressure with respect to all three space variables occur. But the only derivative of the density which occurs is the " total " derivative following a particle of fluid. Now, neglecting conduction and radiation of heat, the entropy is constant for a given particle of fluid, so that this total derivative of density may be replaced by that of the pressure, divided by the square of the velocity of sound. The four equations are then homogenous in the first derivatives of the pressure and the velocities, the coefficients of these derivatives being the velocities and various thermodynamic variables. Hence, in replacing the latter by their values as functions of the pressure on a particular adiabatic, only variations in entropy multiplied by variations in velocity or pressure are neglected in the equations. So, if such products are negligible, it is sufficient to solve the equations of continuity and momentum for the velocity and pressure as if the entropy were constant.

Thus in the case of plane waves the Riemann theory can be used to determine the velocity and pressure. It will perhaps add to the reader's conviction if this fact is deduced independently from the exact characteristic equations for plane waves (Courant and Friedrichs 1948, p. 82):

$$\left.\begin{aligned} du=-\frac{\mathrm{V}-v}{a}dp \quad \text{on} \quad \frac{dx}{dt}=u+a, \qquad du=+\frac{\mathrm{V}-v}{a}dp \quad \text{on} \quad \frac{dx}{dt}=u-a, \\ ds=0 \quad \text{on} \quad \frac{dx}{dt}=u. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (44)$$

If in these equations $(\mathrm{V}-v)/a$ be replaced by the value as a function of p which it takes for one particular value of the entropy, say $(\mathrm{V}-v)/a=\mathrm{G}'(p)$, changes in u and p will be given errors of order the entropy multiplied by those changes themselves. So if variations in u and p are first order and those of entropy are third order, the errors in taking

$$\left.\begin{aligned} u+\mathrm{G}(p)=\text{constant} \quad \text{on} \quad \frac{dx}{dt}=u+a, \\ u-\mathrm{G}(p)=\text{constant} \quad \text{on} \quad \frac{dx}{dt}=u-a, \end{aligned}\right\} \quad . \quad . \quad . \quad (45)$$

(to replace the ordinary Riemann equations) will be of the fourth order. (But if $\mathrm{G}(p)$ were replaced by a function of v which is identical along the adiabatic $s=0$—in fact by the integral (5)—the error would be of the third order.)

Now consider equations (44) at a point directly behind the shock. Here the shock conditions (14) and (17) hold. Given any one of u, p, v, these determine the others. Also at adjacent points on the shock the conditions hold. Hence changes of u along the shock are related in a specific way to changes of p; that is du/dp along $dx/dt=\mathrm{U}$ is a known function deducible from conditions (14) and (17). To be precise, it is

$$\left(\frac{du}{dp}\right)_{\text{shock}}=\frac{(\mathrm{P}-(\partial e/\partial v)_p)v+(\partial e/\partial p)_v p}{2u(\mathrm{P}+\frac{1}{2}p-(\partial e/\partial v)_p)} \quad . \quad . \quad . \quad . \quad . \quad (46)$$

But if $dp/dt=\alpha$ on $dx/dt=u+a$, so that $du/dt=-\alpha(V-v)/a$, and if $dp/dt=\beta$ on $dx/dt=u-a$, so that $du/dt=\beta(V-v)/a$, then

$$\left(\frac{du}{dp}\right)_{\text{shock}}=\frac{(du/dt)_{dx/dt=U}}{(dp/dt)_{dx/dt=U}}$$

$$=\frac{\dfrac{U-u+a}{2a}\left(-\dfrac{\alpha(V-v)}{a}\right)+\dfrac{-U+u+a}{2a}\left(\dfrac{\beta(V-v)}{a}\right)}{\dfrac{U-u+a}{2a}\alpha+\dfrac{-U+u+a}{2a}\beta}$$

$$=\frac{V-v}{a}\,\frac{(a+u-U)\beta-(a-u+U)\alpha}{(a+u-U)\beta+(a-u+U)\alpha}. \qquad (47)$$

This gives the reflection coefficient exactly as

$$\frac{\alpha}{\beta}=\frac{a+u-U}{a-u+U}\left(1-\frac{a}{V-v}\left(\frac{du}{dp}\right)_{\text{shock}}\right)\Big/\left(1+\frac{a}{V-v}\left(\frac{du}{dp}\right)_{\text{shock}}\right) \qquad (48)$$

For a weak shock the numerators in (48) are respectively of the first and second orders in the strength, so that the whole is of the third order. For since, by (16), behind the shock

$$p=P_V v+\tfrac{1}{2}P_{VV}v^2+\tfrac{1}{6}P_{VVV}v^3+P_S\frac{P_{VV}}{12T}v^3+O(v^4), \qquad (49)$$

while by (17) $u=(pv)^{\frac{1}{2}}$, it is easy to deduce that

$$\left(\frac{dp}{du}\right)_{\text{shock}}=\frac{A}{V}\left(1+\frac{P_{VV}}{2P_V}v+\left(\frac{P_{VVV}}{4P_V}+\frac{P_{VV}T_V}{8TP_V}-\frac{5P_{VV}^2}{32P_V^2}\right)v^2+O(v^3)\right), \qquad (50)$$

whereas, by expanding a in powers of v,

$$\frac{a}{V-v}=\frac{A}{V}\left(1+\frac{P_{VV}}{2P_V}v+\left(\frac{P_{VVV}}{4P_V}-\frac{P_{VV}^2}{8P_V^2}\right)v^2+O(v^3)\right). \qquad (51)$$

Hence the reflection coefficient (48), to the third order, using (17) and (11), is

$$\frac{\alpha}{\beta}=\frac{\frac{1}{2}A\sigma_1}{2A}\left(\frac{P_{VV}T_V}{8TP_V}-\frac{P_{VV}^2}{32P_V^2}\right)\left(\frac{2P_V}{P_{VV}}\sigma_1\right)^2\Big/2=\frac{P_V}{64VP_{VV}}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)\sigma_1^3. \qquad (52)$$

The bracket in (52) is recognized as that of (2). Notice, however, that the term in T_V/T due to entropy change has arisen in (52) in a manner different from that in which it appeared in calculating the residual energy. It is due to the second order term in dp/du, not to that in U. This fact seems to the author to reduce still further the likelihood that the agreement of the check on Friedrichs's theory is fortuitous.

Now β, the value of dp/dt on $dx/dt=u-a$, is known approximately from the simple wave theory (on which, however, $\alpha=0$). To a first approximation $p=(2P_V^2/P_{VV})\sigma$; also, by (7), if $dx/dt=u-a$,

$$-A\,dt=(\xi'(\sigma)+At)\,d\sigma+A\,dt, \qquad (53)$$

so that

$$\beta=\frac{dp}{dt}=(2P_V^2/P_{VV})\frac{-2A}{\xi'(\sigma)+At}. \qquad (54)$$

Hence by (52), and since for $\sigma=\sigma_1$, to a first approximation, by (20), the denominator of (54) is $-\frac{1}{2}\sigma_1\,dt/d\sigma_1$,

$$\alpha=\frac{\mathrm{P}_{\mathrm{V}}^3}{8\mathrm{V}\mathrm{P}_{\mathrm{VV}}^2}\left(4\frac{\mathrm{VT}_{\mathrm{V}}}{\mathrm{T}}-\frac{\mathrm{VP}_{\mathrm{VV}}}{\mathrm{P}_{\mathrm{V}}}\right)\sigma_1^2\frac{d\sigma_1}{dt}. \qquad . \quad . \quad . \quad . \quad (55)$$

To see that this value of $(dp/dt)_{dx/dt=u+a}$ is propagated through the simple wave unchanged to a first approximation, consider two adjacent characteristics $dx/dt=u-a$. The time interval Δt between them remains unchanged to a first approximation. At the shock the pressure difference between them, along a characteristic of the other family, is $\alpha\Delta t$, and the velocity difference is $-\mathrm{G}'(p)\alpha\Delta t$ by (45). As p varies on each characteristic, (45) shows that u varies by the change in $\mathrm{G}(p)$, so behind the simple wave (for example)

$$\Delta u-(-\mathrm{G}'(p)\alpha\Delta t)=\mathrm{G}'(p)(\Delta p-\alpha\Delta t). \qquad . \quad . \quad . \quad . \quad (56)$$

But if the differences between the two characteristics are again taken along a characteristic of the other family, then $\Delta u=-\mathrm{G}'(p)\Delta p$. Hence, by (56), $\Delta p=\alpha\Delta t$ as before; that is, behind the simple wave, dp/dt on a characteristic $dx/dt=u+a$ is still given by (55).

Since the shock is, to a first approximation, a characteristic of the family $dx/dt=u+a$, it has been shown that the flow behind the simple wave is the same as if the shock sent out a pulse at time t whose time derivative is (55) and which can therefore differ by at most a constant from expression (42). (Strictly it has only been shown that the waves travelling towards the piston are of this form; but waves travelling in the opposite direction must represent the reflection of the former waves at the piston if the boundary condition there is to be satisfied.)

This checks (42) to within a constant. The constant if present would, by the theory of §8, make the total energy of the fluid tend to infinity with time. But we can also see that it must be absent from the point of view of the present section. For a complete third order solution of the equations of motion and boundary conditions could now be given by specifying the pressure and velocity fields behind the shock as the combination of the simple wave, determined (to the third order) by the motion of the piston, and its third order reflection in the shock (with *its* reflection at the piston). Now the value of the pressure at the piston *immediately* after it first moves may be calculated independently, (i) from the shock conditions, and (ii) from the simple wave theory. The difference must be twice the pressure in the reflected wave at this initial stage (twice because the reflected wave is itself reflected at the piston). The two quantities are in fact the integrals with respect to u, from 0 to u, of (50) and (51) respectively. Their difference is easily verified to be twice (42). This check of (42) at $t=0$, with the previous check of its time derivative, completes the investigation of the accuracy hypothesis concerning the Friedrichs theory. It has been found watertight at every point.

The complete third order solution, which could be obtained in the manner just described, would add a negative pressure distribution to that given by simple wave theory. Now, at large distances from the piston, the reflection of the simple wave at the shock becomes negligible (by (42)), and only the reflection of this reflected wave at the piston modifies the original simple wave. This, travelling in the same direction, turns it into another simple wave (the entropy being effectively uniform at these large distances from the piston), with reduced pressure distribution. This means that the characteristics are slowed down and, therefore, that which now travels with velocity A must be one emitted slightly before the piston comes to rest. This is the one which was taken as the boundary of the simple wave in §5, for the purpose of calculating its energy in §6.

It is now clear that the form of the accuracy hypothesis adopted in §1 was necessary as well as sufficient. For a modification of the simple wave of the kind described really takes place at large distances from the piston, while near it the entropy is modified as stated. It is not surprising that the errors in the Friedrichs theory should be comparable with those obtained by introducing both these modifications, and no others, uniformly.

§ 10. THE DECAYING N-WAVE.

After the careful examination of this single class of problems, each involving only one decaying shock, it is important to discover whether the conclusions for phenomena involving more than one shock are substantially identical or not.

The problem which will be investigated is like that treated in §§1–9, except that the piston is continuously retarded (from its initial impulsively attained positive velocity) through zero speed to a negative velocity (so that it is moving away from the fluid) and is then instantaneously stopped (*not* necessarily in its original position), and subsequently held at rest. The piston, on being instantaneously stopped, emits a second shock, which moves along the tube behind the first while both gradually decay in strength. The motion of the front shock is exactly as in the problem previously studied.

Friedrichs (1948) only treats a special case, but his method is quite general. The flow between the shocks is taken to be the simple wave, at the entropy of the undisturbed fluid, which is compatible with the motion of the piston. The fluid behind the second shock is in its original undisturbed state, and the shock conditions thereat, as at the first shock, are to be satisfied to the second order in the strength. The pressure, after rising discontinuously at the front shock, falls in the simple wave monotonically to below its undisturbed value, to which it then rises discontinuously at the rear shock. This behaviour may be represented graphically as something like an N, whence the wave is called an N-wave.

Considerations as in §1 lead to the formulation of an accuracy hypothesis for the Friedrichs theory of the N-wave, with identical wording to that used in §1. Conclusions as in §§2–9 can then be drawn, and will be displayed in brief, which check the hypothesis and improve the picture of the motion.

But it must be remarked at once that this hypothesis, which is about to be checked, does not forecast that the motion of the rear shock will be given to a similar degree of accuracy by the simple wave theory, at distances from the piston so large that it has decayed to a small fraction of its initial strength, as was indicated for the front shock by the theory of §4. For the front shock, conditions on one side are known exactly. As a consequence, conditions on the other side can be specified, as in § 4, with an error of order the cube of the *local* strength. But matters at the rear shock are different. The conditions behind it are known, on the accuracy hypothesis, with a possible error of order the cube of the *maximum* shock strength. This error communicates itself to the conditions ahead of it, which determine the position of the rear shock in the simple wave; and when the shock has become very weak its motion may therefore be given quite inaccurately by simple wave theory. It will be seen later, however, by use of the improved physical picture, how to find more closely what will happen, and hence, for the time being, the consequences merely of the Friedrichs theory of the N-wave will be set out.

Taking the state behind the rear shock as undisturbed, it is easy to show that the shock conditions in the simple wave thereat have exactly the same form as for the front shock. For example, (19) holds (but notice that σ is negative at the second shock). At time t let the variable σ take the value σ_2 at the second shock. Then the differential equation (21) holds equally with σ_2 for σ_1. But the initial condition is different. If $\sigma=\sigma_3$ at the piston when it comes to rest, say at $t=t_3 \simeq -\xi(\sigma_3)/\mathrm{A}$, then these are the initial values for the second shock, at which therefore (instead of the solution (22))

$$\mathrm{A}t=2\left(\frac{1-2\lambda\sigma_2}{\sigma_2}\right)^2\left[\int_{\sigma_2}^{\sigma_3}\frac{\sigma\,d\xi(\sigma)}{(1-2\lambda\sigma)^3}+\frac{1}{2}\left(\frac{\sigma_3}{1-2\lambda\sigma_3}\right)^2\mathrm{A}t_3\right], \quad . \quad (57)$$

or, rejecting unreliable terms, as in (23),

$$\mathrm{A}t=-\left(\frac{2}{\sigma_2^2}-\frac{8\lambda}{\sigma_2}\right)\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi(\sigma)-\frac{2}{\sigma_2^2}\left(6\lambda\int_{\sigma_3}^{\sigma_2}\sigma^2\,d\xi(\sigma)+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)\right). \quad (58)$$

(But since the expression (19) for U, and hence also the coefficient of $dt/d\sigma_2$ in the differential equation, may contain errors $\mathrm{O}(\sigma_0^3)$, the equation (58) is not reliable for indefinitely small σ_2.)

The work done by the piston before coming to rest is the same as in (26), but with σ_3 replacing the limit of integration 0, and hence with an additional term $\frac{1}{2}\sigma_3^2\xi(\sigma_3)$ inside the square brackets, since in integrating $\sigma\xi$ by parts the integrated term does not vanish at the limit σ_3.

At time t the simple wave is now specified by the inequality $\sigma_2<\sigma<\sigma_1$, *i.e.* it lies between the two shocks. Hence in the expression (28) for its energy the range of integration must be replaced by this one; and the integrated terms $At\frac{1}{2}\sigma_1^2$ and $At\frac{1}{3}\sigma_1^3$ must be replaced by $At(\frac{1}{2}\sigma_1^2-\frac{1}{2}\sigma_2^2)$ and $At(\frac{1}{3}\sigma_1^3-\frac{1}{3}\sigma_2^3)$ respectively. Then, replacing At by its expression (23) or (58), according as it is multiplied by powers of σ_1 or σ_2, the energy in the simple wave becomes

$$\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\Bigg[\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi-4\lambda\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi+6\lambda\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi-4\lambda\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi+6\lambda\int_{\sigma_3}^{\sigma_2}\sigma^2\,d\xi$$
$$+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)+\left(\frac{1}{2}+\frac{2\mathrm{P_V^2}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P_{VV}^2}}\right)$$
$$\times\left(\int_{\sigma_2}^{\sigma_1}\sigma^2\,d\xi+\tfrac{2}{3}\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi+\tfrac{2}{3}\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi\right)\Bigg], \quad . \quad . \quad . \quad . \quad . \quad (59)$$

and the residual energy, obtained by subtracting (59) from the work done by the piston, becomes

$$\mathrm{R}(t)=\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left(\frac{1}{2}+\frac{2\mathrm{P_V^2}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P_{VV}^2}}-6\lambda\right)$$
$$\times\left[\left(\int_{\sigma_1}^{\sigma_0}+\int_{\sigma_3}^{\sigma_2}\right)\sigma^2\,d\xi-\tfrac{2}{3}\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi-\tfrac{2}{3}\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi\right]. \quad . \quad . \quad (60)$$

(Again, (59) and (60) may be inaccurate, since (58) has been used to obtain them, when σ_2 has become too small.)

On the other hand the rate of increase of the total entropy S(t) of the fluid is given by the term (31), due to the gain at the front shock, plus a similar term, obtained by writing σ_2 for σ_1 and changing the sign, due to the gain at the rear shock ; the latter is also positive since $\sigma_2<0$. Integrating with respect to the time t from the initial value $\mathrm{S}(0)=0$, and using the first approximation to (23) or (58) to express dt in terms of $d\sigma_1$ or $d\sigma_2$, the expression

$$\mathrm{S}(t)=\frac{2}{3}\frac{\mathrm{P_V^2}}{\mathrm{VTP_{VV}^2}}\left[6\left(\int_{\sigma_1}^{\sigma_0}+\int_{\sigma_3}^{\sigma_2}\right)\sigma^2\,d\xi-4\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi-4\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi\right], \quad (61)$$

results. Equation (61) is correct to the second order, on the accuracy hypothesis, at all times, since after σ_2 has decreased to a small fraction of its initial value, changes in S(t) due to the rear shock cannot be significant.

Comparison of (61) with (60) shows that when both are true the equation (1) relating the entropy S(t) and the residual energy R(t) still holds for an N-wave. For this period of time the theory of § 8, which is based simply on (1) and the accuracy hypothesis. therefore holds without change for the N-wave. Conditions behind the rear shock are, to the third order, as if at time t a pressure pulse (2) was emitted, in the direction x decreasing, from approximately $x=\mathrm{A}t$, *i.e.* (to within the degree of approximation permitted) from either of the two shocks. In terms of σ, this pulse (2) becomes

$$\frac{\mathrm{P_V^3}}{24\mathrm{VP_{VV}^2}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\sigma_1^3-\frac{\mathrm{P_V^3}}{24\mathrm{VP_{VV}^2}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\sigma_2^3. \quad (62)$$

(Compare (42).)

The first term in (62) corresponds to the reflection of the simple wave at the front shock, as worked out in §9. This reflection is transmitted through the rear shock together with an additional wave (corresponding to the second term in (62)) which may be regarded as the transmission of the simple wave through the rear shock. (The characteristics in the simple wave are gradually caught up by the shock, which, as it does so, touches off a small part of their energy as a transmitted wave travelling in the opposite direction.)

The analysis of the rear shock justifying the last paragraph is as follows. Ahead of it, by §9,

$$\frac{dp}{dt}=\alpha, \quad \text{and} \quad \frac{du}{dt}=-\frac{V-v}{a}\alpha, \quad \text{along} \quad \frac{dx}{dt}=u+a. \qquad (63)$$

But, by (54), immediately ahead of the rear shock,

$$\frac{dp}{dt}=\beta_2=(2P_V^2/P_{VV})\frac{4A}{\sigma_2\, dt/d\sigma_2}, \quad \text{and} \quad \frac{du}{dt}=\frac{V-v}{a}\beta_2, \quad \text{along} \quad \frac{dx}{dt}=u-a. \qquad (64)$$

Behind the rear shock quantities are of the third order and only their first approximation is of interest ; this is given by the theory of sound, on which the shock is a characteristic and, along it, say

$$\frac{dp}{dt}=\delta, \qquad \frac{du}{dt}=-\frac{V}{A}\delta. \qquad (65)$$

Now the pressure *difference* and velocity *difference* across the rear shock are so related that changes in them satisfy (50), since the coefficients in (50) represent the state behind the shock correctly to the second order. Taken with (65) this implies that, ahead of the rear shock,

$$\left[\left(\frac{dp}{dt}\right)_{dx/dt=U}-\delta\right]\Big/\left[\left(\frac{du}{dt}\right)_{dx/dt=U}+\frac{V}{A}\delta\right]=\left(\frac{dp}{du}\right)_{\text{shock}}, \qquad (66)$$

with third order terms neglected on the right and fourth order terms neglected in both numerator and denominator on the left. Substituting in the left-hand side by use of the information in (63) and (64), and maintaining the same degree of approximation, it becomes

$$\left[\frac{U-u+a}{2a}\alpha+\frac{-U+u+a}{2a}\beta_2-\delta\right]\Big/\left[\frac{U-u+a}{2a}\left(-\frac{\alpha(V-v)}{a}\right)\right.$$
$$\left.+\frac{-U+u+a}{2a}\left(\frac{\beta_2(V-v)}{a}\right)+\frac{V}{A}\delta\right]$$
$$=\frac{a}{V-v}\left(\alpha-\delta+\frac{a+u-U}{2A}\beta_2\right)\Big/\left(\frac{a+u-U}{2A}\beta_2-\alpha+\delta\right), \qquad (67)$$

so that

$$\delta=\alpha-\beta_2\frac{a+u-U}{4A}\left(1-\frac{a}{V-v}\left(\frac{du}{dp}\right)_{\text{shock}}\right)$$
$$=\frac{P_V^3}{8VP_{VV}^2}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)\left(\sigma_1^2\frac{d\sigma_1}{dt}-\sigma_2^2\frac{d\sigma_2}{dt}\right), \qquad (68)$$

by (55), (64), (19), (51) and (50).

Integrating (68), it may be deduced that the pressure distribution along the rear side of the rear shock is as in (62). These pressures (along this approximate characteristic) may be taken as propagated acoustically along characteristics of the other family, *i.e.* towards the piston—as has been deduced previously by energy considerations alone.

It may now be seen what are the third order disturbances behind the rear shock which make its equation of motion (58) unreliable after it has greatly weakened. By this time the actual pressure pulses (62) have themselves become weak; it is the *reflection at the piston* of pulses (62) emitted at a much earlier time which are significant. (This reflected wave is given by (41) without the first term in square brackets.) Notice that ultimately, though very gradually, the whole reflected wave catches up the rear shock. For, by (58), the distance of the rear shock behind an imaginary wave emitted by the piston with velocity A on coming to rest, which is practically $-At\sigma_2$, tends to infinity like a small multiple of $t^{\frac{1}{2}}$. So all acoustic waves travelling behind it ultimately, though slowly, catch it up. Their reflection from the shock is negligible, since the reflection coefficient is of the third order. They cannot be transmitted through it, since the time decreases on both characteristics ahead of the shock. Their energy goes into an alteration of the shock path. Thus, since they are rarefaction waves, their incidence on the shock makes it weaken more slowly, and makes the N-wave become wider (on the low pressure side) than on the theory which neglects them. It is only in this indirect way that they transmit their negative energy to the N-wave.

§11. THE ENERGY DISTRIBUTION AFTER A LONG TIME.

This section will treat of the approximate energy distribution, as given by the Friedrichs theory, after the shocks have weakened to a small fraction of their initial strengths, and also of the subsequent modifications due to the gradual overtaking of the rear shock by the pressure wave behind it.

The energy distribution on the Friedrichs theory, as $t \to \infty$, will be divided into that remaining permanently in the N-wave (whose width increases as its amplitude decreases), that due to increase in entropy of the fluid through which the wave has passed (with, ultimately, no change in pressure), and that in the pressure wave behind the rear shock. It is only after these have become almost stationary at their limiting values that the gradual process of transfer of energy from the third form to the first becomes important. For this reason the said limiting values will first be determined.

The expressions which have been obtained for all these energies with t arbitrary are simplified by letting $t \to \infty$. For then, by (23) and (58), σ_1 and σ_2 tend to 0 (naturally, because the shock strengths tend to 0). The energy (59) in the simple wave becomes

$$\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left[\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi+6\lambda\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)\right]. \quad . \quad . \quad . \quad (69)$$

The energy due to increased entropy, after the original pressure has been recovered, by (34) and (61), is

$$T\left(1-\frac{T_V P}{TP_V}\right)S(\infty)=\frac{4P_V^3}{VP_{VV}^2}\left(1-\frac{T_V P}{TP_V}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi, \quad . \quad . \quad . \quad (70)$$

and the remainder of the residual energy, corresponding to the pressure wave behind the rear shock, is

$$T\left(\frac{1}{2}\frac{T_V P}{TP_V}-\frac{1}{8}\frac{PP_{VV}}{P_V^2}\right)S(\infty)=\frac{P_V^2 P}{2V^2P_{VV}^2}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi \quad . \quad (71)$$

by (1). For the problem of §§1–9, without a rear shock, (69), (70) and (71) are correct at all sufficiently large times. For that of §10 the gradual transfer of the energy (71) to the N-wave will make the energy of the N-wave, after a very long time indeed, equal to the sum of (69) and (71), namely to

$$\frac{2PP_V}{VP_{VV}}\left[\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi+\left(\tfrac{1}{2}+\frac{P_V}{P_{VV}}\left(\frac{3}{V}-\frac{P_{VVV}}{P_V}+2\frac{T_V}{T}\right)\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)\right]. \quad . \quad . \quad . \quad (72)$$

The energy (70) is positive for all fluids; the energy (71) is probably negative for all fluids (§8), though zero to the third order for monatomic gases. The energy (69) may be positive or negative, and normally the first term in the square brackets therein will be of a higher order of magnitude than the others, and will dominate this part of the energy and make it more important than the residual energy (70) and (71). We have then an N-wave with one "upright" of the N always taller than the other, and the internal energy on that taller side of the wave due to greater volume changes on that side predominates in importance at all times.

But there is one problem, of special interest, in which the first term in (69) does not dominate the others. This is the case when the piston is stopped precisely on reaching its initial position. It is interesting as being directly analogous to the problem of the steady two-dimensional supersonic flow past an aerofoil (neglecting boundary layer effects). The condition is that over the path of the piston

$$\int_{\sigma_0}^{\sigma_3}u(dt/d\sigma)\,d\sigma=0,$$

whence, by formulæ of §5,

$$\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi+\left(\frac{1}{2}+\frac{3P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)=0. \quad . \quad . \quad . \quad (73)$$

Then the energy (69) in the N-wave *, on substituting for λ from (19), becomes

$$\frac{2PP_V}{VP_{VV}}\left(\frac{1}{4}+\frac{P_VT_V}{P_{VV}T}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (74)$$

---

* This energy is still due to an inequality in the "uprights" of the N, though a smaller one than when the first term in (68) dominates the others.

Thus, if the piston comes to rest in its initial position, the approximate distribution of energy, when the shocks have decayed to a small fraction of their initial strengths, as between the N-wave, the increased entropy of the fluid as a whole, and the pressure waves behind the N-wave, is in the ratios (by (74), (70) and (71))

$$\left(\frac{1}{2}\frac{\mathrm{T_V P}}{\mathrm{TP_V}}+\frac{1}{8}\frac{\mathrm{PP_{VV}}}{\mathrm{P_V^2}}\right):\left(1-\frac{\mathrm{T_V P}}{\mathrm{TP_V}}\right):\left(\frac{1}{2}\frac{\mathrm{T_V P}}{\mathrm{TP_V}}-\frac{1}{8}\frac{\mathrm{PP_{VV}}}{\mathrm{P_V^2}}\right). \quad . \qquad (75)$$

These depend only on the physical properties of the fluid.

For a perfect gas with constant adiabatic index these ratios become

$$\frac{5\gamma-3}{8\gamma}:\frac{1}{\gamma}:\frac{3\gamma-5}{8\gamma}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (76)$$

and for air they are

$$\frac{5}{14}:\frac{10}{14}:-\frac{1}{14}\,^{*}.$$

Notice that in this one case the total energy of the fluid is $\mathrm{TS}(\infty)$. Thus, after $\mathrm{S}(t)$ has reached its limiting value, the energy is that associated with the total entropy change when volume changes are small and integrate out to zero. That the integrated specific volume changes should be zero is inevitable, to a first approximation, since the space occupied by the fluid is unchanged if the piston returns to its initial position. Notice however that the effect is achieved by cancellation of the decreased volume in the N-wave with the increased volume (due to increased entropy without change of pressure) near the piston. Observe also that, before $\mathrm{S}(t)\simeq\mathrm{S}(\infty)$, the volume changes in the N-wave are too large, compared with the entropy changes, to be treated linearly (and in fact the energy exceeds $\mathrm{TS}(t)$, being $\mathrm{TS}(\infty)$ at all times after the piston has come to rest).

As time goes on the energy in the third form is transferred to that in the first. Ultimately (one may suppose) in all cases, and not merely approximately, all the energy which has not gone into increasing the entropy of the fluid without change of pressure is in progress to infinity with the N-wave, after that portion which has been released and reflected at the piston has once more overtaken it. (It seems unlikely that, even theoretically and after vast times, this plane N-wave would be damped by viscosity or heat conduction. For it follows from (7) that $d\sigma/dx\sim(\mathrm{A}t)^{-1}$ as $t\to\infty$, so that the gradients of all physical quantities have an approximately uniform value in the N-wave, decreasing as the reciprocal of the time. The energy dissipation in all time might then be expected to be bounded by a negligible integral of the form $\mathrm{O}\left(\int^{\infty}\mu t^{-2}\,.\,t^{\frac{1}{2}}\,dt\right)$, where μ is the viscosity).

---

* For a liquid at ordinary pressures they are practically 0 : 1 : 0, owing to the low value of $\kappa\mathrm{P}=\mathrm{P/VPv}$. (But if the piston does not return to its original position, the energy in the N-wave still predominates.)

The path of the rear shock, as it gradually moves behind the path given by the Friedrichs theory, thus increasing the width of the N-wave on the low pressure side to permit the negative energy (71) to pass into the N-wave, would be calculable to some extent, if necessary. The errors in (18), as a formula for giving the conditions at the rear shock, if the flow is disturbed behind it, are due, to a first approximation, to taking A incorrectly, and to taking U and v not relative to the velocity and volume behind the shock. Thus, if suffix 1 refers to conditions behind the rear shock, the error in (18) is $u_1+a_1-A(1+P_{VV}v_1/2P_V)$. When the shocks have weakened considerably, so that s is practically zero behind them, this expression becomes u_1+Vp_1/A approximately, a quantity constant along characteristics $dx=A\,dt$, and equal by (41) to

$$\frac{VT}{8A^2}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)S'\left(\frac{1}{2}\left(t-\frac{x}{A}\right)\right). \quad . \quad . \quad . \quad . \quad (77)$$

Adding this to (19), we see that the differential equation (21) (with σ_2 for σ_1) holds with (77) subtracted from the coefficient of $A\,dt/d\sigma_2$. This equation, in which $t-x/A$ is practically $-t\sigma_2$, could be treated approximately, but this will not be attempted here since the physical nature of the process is already clear, and the purposes of the paper have all been accomplished.

Some idea of the time scale of the final energy transfer is however desirable. The pulse (62) will still be considerable when σ_1 or σ_2 is, say, half its initial value, so that by (23) or (58) t is of order $\max|\xi|/A\sigma_0 \simeq t_3/\sigma_0$. Its reflection in the piston catches up with the rear shock when, at it, $\frac{1}{2}(t-x/A)$, *i. e.* $-\frac{1}{2}t\sigma_2$, is of this order, so that by (58)

$$\frac{t_3\sigma_0}{\sigma_2}\approx\frac{t_3}{\sigma_0}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (78)$$

Thus, when the transfer of energy from the rear pressure wave to the N-wave is in full swing, the rear shock strength is of order the square of its initial value, and the time t is of order the duration of motion of the piston t_3, divided by the cube of the initial shock strength.

The ultimate asymptotic behaviour of σ_2 (which measures the rear shock strength, and also its position, at time t) may also be obtained without difficulty. After a long time, when disturbances have become very small, only the Pv term in the specific energy affects the integrated energy in the N-wave, and this is therefore $(PP_V/VP_{VV})(\sigma_1^2-\sigma_2^2)At$. Equating this expression—with $\sigma_1^2\,At$ replaced by its limiting form

$$\int_0^{\sigma_0}(2\sigma+12\lambda\sigma^2)\,d\xi$$

as given by (23)—to expression (72) gives that (neglecting σ_0^3)

$$\lim_{t\to\infty}(At\sigma_2^2)=-2\int_{\sigma_3}^{0}\sigma\,d\xi-12\lambda\int_{\sigma_3}^{0}\sigma^2\,d\xi-\sigma_3^2\xi(\sigma_3)+\left(\frac{1}{2}-\frac{2P_VT_V}{P_{VV}T}\right)\int_{\sigma_3}^{0}\sigma^2\,d\xi. \quad . \quad . \quad . \quad (79)$$

The Friedrichs theory gives the limit as (79) without its last term. Thus that theory describes the final stage of decay of the rear shock correctly only to a first approximation, although it describes that of the front shock correctly to a second approximation.

REFERENCES.

BRINKLEY, S. R., and KIRKWOOD, J. G., 1947, *Phys. Rev.*, **71**, 606–611.

BRODERICK, J. B., 1949, *Q.J.M.A.M.*, **2**, 98–120.

CASSEN, B., and STANTON, J., 1947, *Bull. Amer. Phys. Soc.*, **22**.

COURANT, R., and FRIEDRICHS, K. O., 1948, *Supersonic Flow and Shock Waves* (New York : Interscience Publishers, Inc.).

DUMOND, J. W. M., COHEN, E. R., PANOVSKY, W. K. H., and DEEDS, E., 1946, *J. Acoust. Soc. Amer.*, **18**, 97–118.

FRIEDRICHS, K. O., 1948, *Comm. App. Maths.*, **1**, 211–245.

LIGHTHILL, M. J., 1944, *A.R.C. R. & M.* 1930 ; 1948, *Q.J.M.A.M.*, **1**, 76–89 ; 1949 a, *Phil. Mag.*, **40**, 214–220 ; 1949 b, *Proc. Roy. Soc.* A, **198**, 454–470 ; 1950, *Ibid.*, **200**, 554–565.

MARTIN, M. H., 1947, *Q. App. Math.*, **4**, 330–348.

TAYLOR, G. I., 1950, *Proc. Roy. Soc.* A, **201**, 159–186.

WHITHAM, G. B., 1950, *Proc. Roy. Soc.* A, **201**, 89–109 ; *The Propagation of Spherical Blast* (unpublished).

XCIX. *Energy Losses of Ionizing Particles at Relativistic Velocities in the Photographic Plate.*

By H. Messel and D. M. Ritson,
Dublin Institute for Advanced Studies*.

[Received June, 26, 1950.]

The theory of ionization losses at relativistic velocities in condensed media was first treated quantitatively by Fermi (1940). Fermi showed the importance of considering the dielectric properties of the medium, and his calculations were later extended by Halpern and Hall (1948) to cover a rather general model for the dielectric properties. This theory is based on well established laws of molecular physics and classical electrodynamics, and the presence of any large deviations from it would be most surprising. In recent investigations Halpern and Hall (1948), Bradt, Kaplan and Peters (1950,) Occhialini (1950), with the photographic plate on the grain density of relativistic tracks, no increases of grain density greater than 10 per cent have been observed up to momenta of the order of 10^4 Mc. These observations have been interpreted as due either to a failure of the Halpern–Hall theory or to the hypothesis that ionization caused by distant collisions will not contribute to the development of Silver Bromide grains Occhialini (1950), Powell (1950), Wilson (1950). Both of these assumptions are difficult to accept.

However, detailed examination of the problem shows that these interpretations are based on an incorrect application of the results given by Halpern and Hall. Halpern and Hall were interested in the energy losses of particles traversing large thicknesses of matter and therefore calculated the total loss of energy in all collision processes.

The quantity required for calculations on the photographic plate as also for certain applications in cloud chamber measurements is the " probable ionization ". The development of a Silver Bromide grain will be a function of the energy actually converted into ionization losses in the grain, and will not be a function of the total energy loss. δ-rays with ranges very much greater than the radius of the grains will lose only a small fraction of their energy in the grain of Silver Bromide, and thus the energy losses, in the form of such δ-rays, of an ionizing particle will not contribute to the development of the Silver Bromide grain. The quantity required is the energy lost in all collision processes involving energies less than some maximum energy value $E_{\max}$. This quantity is defined as the " probable ionization ".

The Bethe–Bloch formula for " probable ionization " losses is as follows :

$$-\frac{dW}{dx}=\frac{2\pi ne^4}{mv^2}\left\{\log\frac{mv^2E_{\max}}{I^2Z^2(1-\beta^2)}+(1-\beta^2)\right\},$$

where all the symbols have their usual meaning.

---

* Communicated by Professor L. Jánossy.

For large values of the momentum the theory of Halpern and Hall shows the Bethe–Bloch theory overestimates the energy loss by an amount Δ given by

$$\Delta=\frac{2\pi ne^4}{mc^2}\left\{\log\frac{ne^2h^2}{\pi m\mathrm{I}^2\mathrm{Z}^2(1-\beta^2)}-1\right\},$$

thus giving an asymptotic form for the " probable " energy loss for high momenta of

$$-\frac{d\mathrm{W}}{dx}=\frac{2\pi ne^4}{mc^2}\left\{\log\frac{\pi m^2c^2\mathrm{E}_{\max}}{ne^2h^2}+1\right\}.$$

This result is identical with that given by Fermi (1940) and is quite independent of any of the assumptions as to the molecular structure of the medium.

For low momenta in the absence of conduction electrons the Halpern and Hall correction is small and thus the ratio of the " probable ionization " losses is insensitive to the minimum value for the "probable ionization" losses to any assumptions made about the molecular structure.

On detailed examination of the Halpern and Hall theory we find that the relativistic increase is entirely associated with the term related to "Cerenkov Radiation ". We have investigated the possibility that energy loss in the form of Cerenkov Radiation will be able to escape from the crystals. This contribution is given by the second term in the formula (28) of the Halpern and Hall paper. The term is

$$\frac{2\pi e^2}{v^2}\int_0^\infty -\theta v\left(\frac{1}{\epsilon}+\frac{1}{\epsilon^*}-2\beta^2\right)dv,$$

θ is a function which is 0 for $(1-\beta^2\epsilon)>0$ and π for $(1-\beta^2\epsilon)<0$. On a multifrequency theory for a dielectric ϵ will be large in the neighbourhood of a resonance frequency and less than unity or small elsewhere. Thus the contributions to the Cerenkov Radiation are mainly from frequencies in the neighbourhood of resonance lines. There will be some loss of " Cerenkov Radiation " particularly for frequencies associated with the K and L levels for which the mass absorption coefficient will be less than the order of 2000 g./cm.^{2} necessary for stopping the radiation from leaving the crystals of Silver Bromide. The theory is thus only approximate.

Fig. 1 shows a graph of the ratio of " probable ionization " to minimum " probable " ionization. For comparison purposes the dotted line shows the ratio of total ionization to minimum total ionization. The difference is due entirely to the contribution from high energy δ-rays. $\mathrm{E}_{\max}$ was taken as 5 kV. corresponding to a range of 0·2μ in Silver Bromide. We used the results of Halpern and Hall, calculated for iron. This is a fair approximation as we were concerned only with the ratios of the ionization. We note that the total relativistic increase is only 15 per cent. The reason for the difference between this result and that actually observed in a cloud chamber gas is due quite simply to the fact that the onset of the polarization effects occur at much lower momenta due to the dielectric properties of the medium. In air we should expect the polarization effects to only assume importance at momenta of the order of one hundred times greater than those for solid Silver Bromide. Thus we should expect an increase of probable ionization logarithmically up to the order of 1·7, and beyond that the ionization should become constant.

Fig. 1.

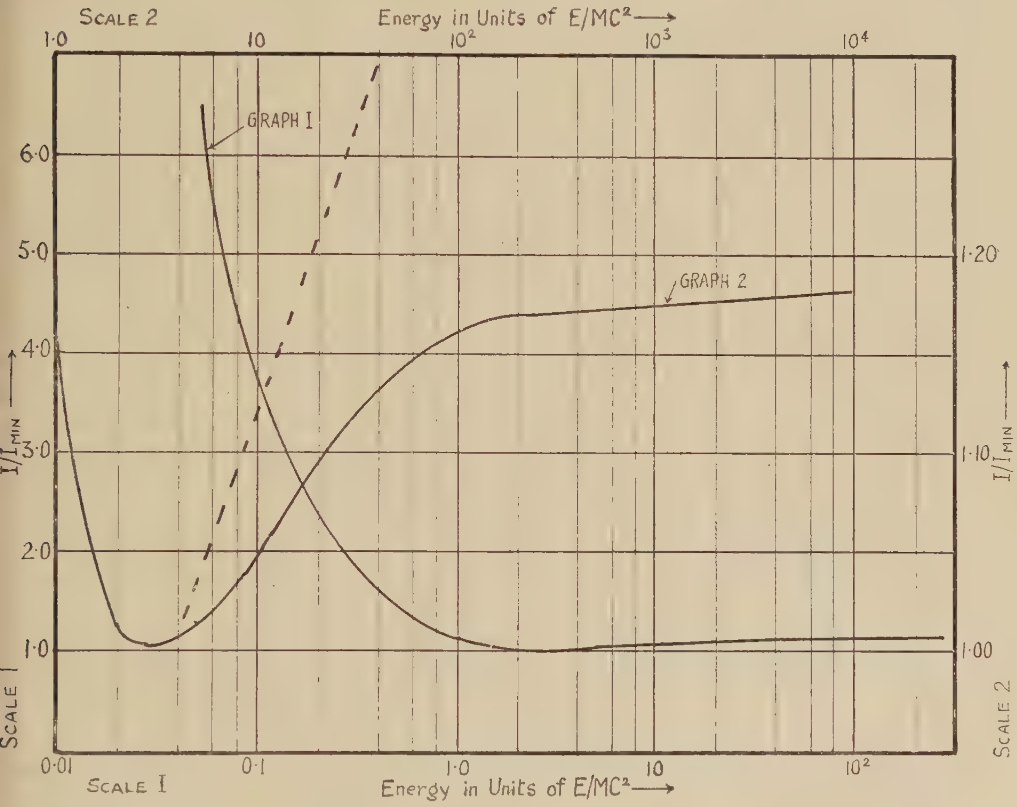


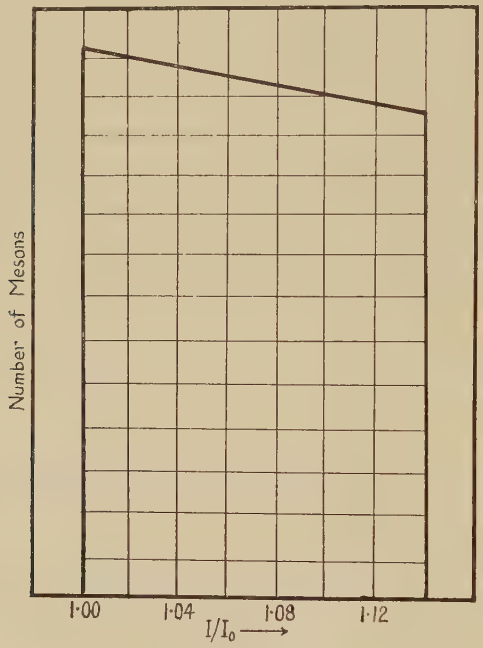
Fig. 2 shows the histogram of the ionization versus numbers calculated for the sea-level spectrum of Cosmic Rays given by Wilson (1946). We note that it is a broad distribution from 1 to 1·15 I_{min} and the mid point is at 1·07 I_{min}. The usual method of determining minimum ionization is to plot such a histogram and determine the mid point, and take this as corresponding to minimum ionization (Brown, Camerini, Fowler, Muirhead, Powell and Ritson 1949). Such a method, as can be seen by reference to fig. 11, will however determine a value of "minimum ionization" of the order of 1.07 I_{min}. The results for high altitudes are not likely to be very different. The energy spectra will contain more low energy mesons near minimum ionization, but this will be counterbalanced by more electrons with ionizations near 1·15 I_{min}.

With measurements as made at present the maximum possible rise in ionization thus comes out at about 7 per cent above the experimentally determined minimum.

Present available data gives no decisive evidence either for or against the above relativistic increase. It would however be important if it could be established with certainty. It would then provide discrimination between mesons and protons in the 10^9eV. range in measurements with the photographic plate.

We are indebted to Professor L. Jánossy and Dr. J. G. Wilson for discussions on this problem.

Fig. 2.



REFERENCES

BRADT, KAPLAN and PETERS, 1950, *Helv. Phys. Acta*, **23**, 24.

BROWN, CAMERINI, FOWLER, MUIRHEAD, POWELL and RITSON, 1949, *Nature*, **163**, 47.

CCCHIALINI, G. P. S., 1950, *Bulletin du Cectre de Physique Nucléaire Bruxelles*, No. 11.

FERMI, E. 1940, *Phys. Rev.*, **57**, 485.

HALPERN, O, and HILL, H, 1948, *Phys Rev.*, **73**, 477.

POWELL, C. F., 1950, Edinburgh Conference, *Nature*, **165**, 54.

WILSON, J. G., 1946, *Nature*, **158**, 144.

WILSON, J. C., 1950, Edinburgh Conference, *Nature*, **165**, 54.

C. *A Theory of the Plastic Bulging of a Metal Diaphragm by Lateral Pressure.*

By R. Hill, M.A., Ph.D.,
Metal Flow Research Laboratory, Sheffield*.

[Received July 27, 1950.]

Summary.

Explicit formulæ are obtained for the stresses in a metal diaphragm which is bulged plastically by lateral pressure. The predicted influence of work-hardening on the shape of the profile, and on the relation between polar strain and curvature, agrees well with experimental data. A simple expression is developed for the instability strain.

§1. Introduction.

The ductility of sheet metal under biaxial stress is often examined by means of the so-called " bulge test ". A uniform plane sheet is placed over a die with a circular or elliptical aperture and is firmly clamped around the perimeter. An increasing hydrostatic pressure is applied to one side of the diaphragm, causing it to bulge through the aperture. From the measured profile and thickness of the plastically-deformed diaphragm near the pole it is possible to calculate the local state of stress in terms of the applied pressure. If, in addition, the state of strain is measured by means of a scribed grid, the stress–strain characteristics of the metal under biaxial tension are obtained. An especial advantage of this test over any other simple one is that a greater range of pre-instability strain can be attained.

To understand the significance of this test it is desirable to have a theory which explains how the varying shape of the bulge, and the distribution of strain, depend on the properties of the metal. The theory should also be able to predict the moment when the process becomes unstable. The only published work of importance appears to be that of Gleyzal (1948) whose analysis is, however, restricted to small strains. Gleyzal formulated the necessary differential equations for a circular aperture and integrated them numerically in a particular case, but did not obtain a general solution. Moreover, he adopted physically unacceptable stress–strain relations in which the components of total strain are made to depend uniquely on the current stress. The present investigation is an attempt to obtain an explicit solution, valid for large deformations and based on the Lévy–Mises stress-strain equations.

---

* Communicated by the Author.

§2. BASIC EQUATIONS.

Consider a circular diaphragm of radius a and thickness t_0, where t_0/a is so small that bending and shearing stresses can be disregarded. Let h be the polar height of the bulge when the applied pressure is p. The bulge being a surface of revolution, we can take cylindrical coordinates (r, z) for a particle, referred to axes in and perpendicular to the original plane (fig. 1). The initial radius to a particle is denoted by s, the current local thickness by t, and the inclination of the surface to the plane of the edge by θ.

By well-known geometrical formulæ the principal radii of curvature are

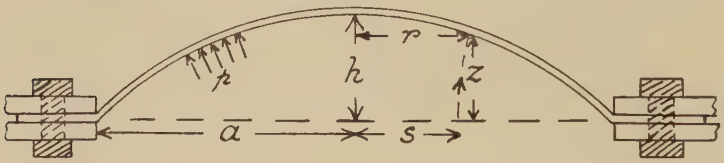
$$\rho_c = r \operatorname{cosec} \theta, \qquad \rho_t = \sec \theta \frac{dr}{d\theta}, \qquad . \ . \ . \ . \ . \ . \ . \quad (1)$$

where ρ_t refers to a section such as fig. 1, and ρ_c to a perpendicular section containing the surface normal. The radii are not independent since the elimination of θ from (1) leads to

$$\frac{r}{\rho_c}\frac{d\rho_c}{dr} = 1 - \frac{\rho_c}{\rho_t}. \qquad . \ . \ . \ . \ . \ . \ . \ . \ . \quad (2)$$

The radii are equal at the pole, but generally nowhere else unless the bulge happens to be a spherical cap.

Fig. 1.



Let σ_c and σ_t be the principal components of stress in the surface (the suffix c denotes the circumferential or hoop direction, and the suffix t denotes the tangential direction in the plane of fig. 1). The equations of equilibrium for a thin membrane in directions normal and tangential to the surface are respectively

$$\frac{p}{t} = \frac{\sigma_c}{\rho_c} + \frac{\sigma_t}{\rho_t}, \qquad \frac{\partial}{\partial r}(rt\sigma_t) = t\sigma_c. \qquad . \ . \ . \ . \ . \ . \ . \quad (3)$$

Alternatively, these may be expressed by means of (1) and (2) as

$$\left.\begin{aligned} \sigma_c &= \frac{p}{2t}\frac{\partial}{\partial r}(r\rho_c) = \frac{p}{2t}\left(2 - \frac{\rho_c}{\rho_t}\right)\rho_c, \\ \sigma_t &= \frac{p}{2t}\rho_c \qquad\quad = \frac{p}{2t} r \operatorname{cosec} \theta. \end{aligned}\right\} \qquad . \ . \ . \ . \ . \ . \ . \quad (4)$$

The formula for σ_t may also be obtained by considering the equilibrium of any cap of the bulge. The stresses are expressed by (4) directly in terms of the pressure and the bulge geometry. Notice that if $\rho_c \gtrless \rho_t$ then $\sigma_c \lessgtr \sigma_t$.

The yield criterion is taken to be

$$f(\sigma_c, \sigma_t)=\sigma(\epsilon), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where $\sigma(\epsilon)$ is the stress–strain relation in uniaxial tension or compression, ϵ is the equivalent logarithmic strain (to be defined later), and f is a homogeneous function of the stresses which does not vary with the amount of deformation.

We neglect elastic strains and adopt the Lévy–Mises relations

$$\frac{\dot{\epsilon}_c}{\dot{\epsilon}_t}=\frac{2\sigma_c-\sigma_t}{2\sigma_t-\sigma_c}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $\dot{\epsilon}_c$ and $\dot{\epsilon}_t$ are the principal strain-rates in the surface. The condition for incompressibility is

$$\dot{\epsilon}_t+\dot{\epsilon}_c+\frac{\dot{t}}{t}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

If $u=r-s$ is the radial displacement of a particle, and $v=\dot{u}$ is its radial speed, then

$$\dot{\epsilon}_c=\frac{v}{r}, \quad \dot{\epsilon}_t=\frac{\partial v}{\partial r}+\dot{\theta}\tan\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

For convenience, rates of change are measured with respect to the height h, which is a suitable monotonically varying quantity. The derivative following a particle is then

$$\frac{\partial}{\partial h}+v\frac{\partial}{dr},$$

so that, for example,

$$v=\left(\frac{\partial}{\partial h}+v\frac{\partial}{\partial r}\right)u \quad \text{or} \quad v=\frac{\partial u}{\partial h}\Big/\left(1-\frac{\partial u}{dr}\right).$$

The boundary conditions on the clamped edge are

$$u, v, z=0 \text{ when } r=a, \text{ for all } h. \quad . \quad . \quad . \quad . \quad (9)$$

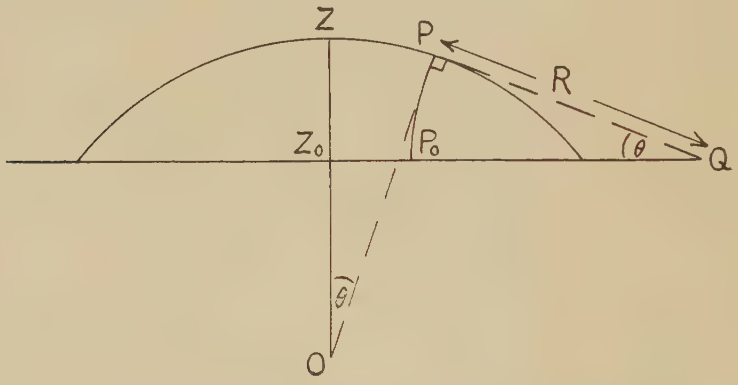
In practice the die is provided with a rounded edge to avoid premature fracture there, and the sheet is clamped around a circle of radius somewhat greater than a. The assumed boundary condition is therefore not strictly satisfied, but, when the radius of the rounded die is less than about $a/10$, the discrepancy between theory and experiment due to this cause appears to be negligible.

The condition (9) implies that the hoop strain vanishes at the edge. We shall see later that the tangential strain may or may not be zero at the edge, depending on the particular material. When the metal is thoroughly annealed, with a vanishingly small yield-stress, it is evident that the edge must strain and work-harden in order that a sufficient tangential stress is available to support the thrust of the pressure over the whole bulge. Hence, according to (6), $\sigma_c=\frac{1}{2}\sigma_t$ at the edge. The edge may, however, cease to deform later when it has become sufficiently hardened in relation to the rest of the diaphragm. On the other hand, when the metal is cold-worked and has a sharp yield-point, σ_c is not necessarily equal to $\frac{1}{2}\sigma_t$ at the edge, and deformation there is then impossible (elastic strains being disregarded). In particular, when plastic bulging is about to begin, both principal stresses are everywhere equal to the initial yield stress. The absence of peripheral strain has been observed in hard copper by Brown and Sachs (1948).

§3. A SPECIAL SOLUTION.

Before trying the general problem it is worthwhile examining whether there exists a particularly simple solution for some special stress–strain curve. Let us see what follows from the trial hypothesis that the particles describe circular paths which are, moreover, orthogonal to the momentary profile (fig. 2). Let R be the radius of the path of the particle initially at radius s. The *conventional* circumferential and tangential strains are

Fig. 2.



$$e_c=(1-\cos\theta)\frac{\mathrm{R}}{s},\quad e_t=-(1-\cos\theta)\left(1+\frac{d\mathrm{R}}{ds}\right).$$

Assuming further that $e_c=e_t$, or

$$\frac{d\mathrm{R}}{ds}+\frac{\mathrm{R}}{s}+1=0,$$

we find that

$$\mathrm{R}=\tfrac{1}{2}\left(\frac{a^2}{s}-s\right). \qquad (10)$$

since $\mathrm{R}=0$ at the edge for zero hoop strain. This equation, being equivalent to $s(2\mathrm{R}+s)=a^2$, is the condition that the particle paths are arcs of coaxal circles having the edges as limiting points. The successive profiles of the diaphragm are therefore the orthogonal set of circular arcs passing through the limiting points. It follows that the diaphragm is deformed into a spherical cap of radius

$$\rho=\tfrac{1}{2}\left(\frac{a^2}{h}+h\right). \qquad (11)$$

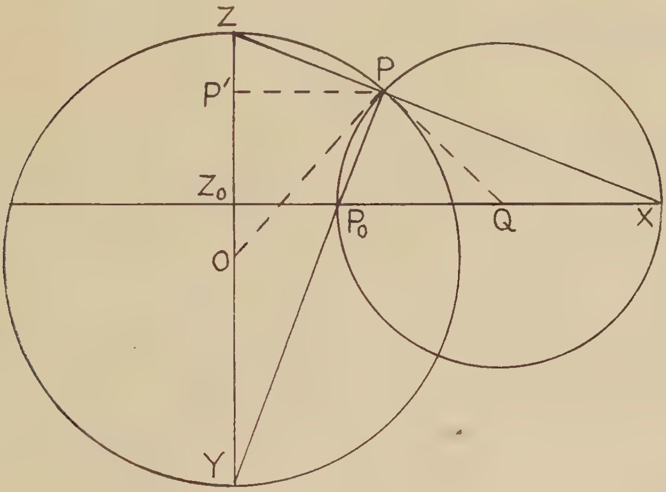
Since the principal strains in the surface are equal, the distortion of every element of the sheet is a uniaxial compression normal to the surface. The equivalent logarithmic strain ϵ is therefore $\ln(t_0/t)$. Moreover, whatever the stress–strain relations, and provided only that the material is isotropic, $\sigma_c=\sigma_t$. That is, the state of stress in every element is a balanced biaxial tension. This is equivalent to a uniaxial compressive stress if the hydrostatic component of stress is without influence on yielding. Hence (5) must reduce to

$$\sigma_c=\sigma_t=\sigma(\epsilon), \text{ where } \epsilon=\ln(t_0/t).$$

From the first equation of (3), $p=2\sigma t/\rho$ and hence the product has the same value throughout the diaphragm at each stage. This demands that the stress–strain curve is

$$\sigma=\sigma_0 \exp \epsilon, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Fig. 3.



where σ_0 is the initial yield stress. The second equation of (3) is now satisfied identically, and the formula for the pressure becomes

$$p=2\sigma_0 t_0/\rho=4\sigma_0 t_0 h/(a^2+h^2). \qquad . \quad . \quad . \quad . \quad . \quad (13)$$

The pressure reaches its maximum when the bulge is a hemisphere. Our initial hypotheses have thus been proved consistent and we have obtained the solution for a special material. Its stress–strain curve is, however, unlike that of any actual metal in that the rate of hardening rises with increasing strain. In fact, the material extends under a constant load. None the less, the relation between the polar strain and curvature is in good agreement with that measured for various metals, as will now be seen.

We first obtain an expression for the strain at any point in terms of the polar height. Consider any two circles with centres O and Q (fig. 3) and any line through one of their points of intersection P meeting the circles in Z and X. Let ZOY and XQP_0 be the respective diameters through Z and X. Then the points P, P_0, and Y are collinear since angles ZPY and XPP_0 are both right-angles. Let Z_0 be the meet of the diameters XP_0 and ZY. Then, if the circles cut orthogonally, so do these diameters. For angle XZ_0Y is the sum of angles OZP and QXP, and therefore of angles OPZ and QPX ; the sum of the latter is a right-angle since the radii OP and QP cut orthogonally, by hypothesis. Translating these results in terms of the geometry of the diaphragm, we see that if Z is the pole of the bulge and P any particle (Z_0 and P_0 being their initial positions), then PP_0 produced passes through Y, the other end of the diameter through Z. To find the final position of a particle initially at P_0, therefore, it is only necessary to produce YP_0 to cut the profile of the diaphragm.

If P′ is the foot of the perpendicular from P to OZ, the conventional hoop strain is

$$\frac{PP'}{Z_0P_0}-1=\frac{YP'}{YZ_0}-1=\frac{Z_0P'}{YZ_0},$$

by similar triangles. Hence, since $Z_0P'=z$ and $YZ_0=a^2/h$,

$$e_c=e_t=\frac{hz}{a^2}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The relation between the curvature and the through-thickness logarithmic strain at the pole can therefore be expressed, parametrically in terms of the polar height, as

$$\epsilon=2\ln\left(1+\frac{h^2}{a^2}\right), \quad \frac{\rho}{a}=\left(1+\frac{h^2}{a^2}\right)\Big/\frac{2h}{a}. \qquad . \quad . \quad . \quad (15)$$

Brown and Thompson (1949) found that the relations between polar curvature and strain for annealed stainless steel, annealed low-carbon steel, and certain aluminium alloys (both hard and soft), differ only slightly. The relation (15) is in excellent agreement with their mean curve up to a strain of order 0·4. It is only at greater strains that a discrepancy occurs, due partly to the occurrence of an instability and partly to the increasing dissimilarity of the stress–strain curve (12) from any actual one. The reason for the insensitivity of the $(\rho/a, \epsilon)$ relation to variations in the properties of the metal will become clear when the general solution is considered.

§4. The Instability Strain.

Brown and Sachs found that when the pressure reaches its maximum a secondary bulge begins to form over about the top half of the diaphragm, and that this continues under a falling pressure. In this sense the pressure maximum can be regarded as the moment when the process becomes unstable. We can calculate the corresponding instability strain by means of (15), which we have seen to be a good approximation for a wide variety of metals. The relation between the pressure and the polar curvature is

$$p=2\sigma t/\rho \quad \text{where} \quad \epsilon=\ln(t_0/t),$$

since the local stress-state is a balanced biaxial tension equal to the yield stress σ. The pressure is a maximum when

$$\frac{1}{\sigma}\frac{d\sigma}{d\epsilon}=1+\frac{1}{\rho}\frac{d\rho}{d\epsilon}, \quad \dots\dots\dots \quad (16)$$

or, from (15), when

$$\frac{1}{\sigma}\frac{d\sigma}{d\epsilon}=\frac{3}{2}-\frac{\rho}{2h}.$$

On expanding in powers of ϵ, this reduces to

$$\frac{1}{\sigma}\frac{d\sigma}{d\epsilon}\simeq\frac{11}{8}-\frac{1}{2\epsilon}, \quad \dots\dots\dots \quad (17)$$

with an error of less than 1 per cent in the relevant range. If the stress–strain curve can be fitted by a power law $\sigma=A\epsilon^n$, the polar strain at maximum pressure is

$$\epsilon_m=\frac{4}{11}(2n+1). \quad \dots\dots\dots \quad (18)$$

For greatest accuracy the exponent n should be chosen to give best fit to the actual stress–strain curve in the neighbourhood of the instability strain.

In uniaxial tension the strain at maximum load is equal to n; in a thin spherical shell expanded by internal pressure, to $2n/3$; and in a plane sheet under balanced biaxial tension, to $2n$. In all these tests the instability strain is vanishingly small for a heavily cold-worked metal, but is equal to 4/11 in the bulge test. This strikingly demonstrates the superiority of the latter as a means of obtaining stress–strain characteristics.

Lankford and Saibel (1947) have attempted to calculate the instability strain in the bulge test by assuming (without previous analysis) that the bulge is spherical and that the strain is given by (14), but with logarithmic strain instead of conventional strain. This therefore agrees with the present analysis only when the deformation is small, which is certainly not so at maximum pressure. Lankford and Saibel's final result is still further vitiated by a gross error committed in taking partial derivatives.

§ 5. GENERAL SOLUTION FOR COLD-WORKED METALS.

The stress–strain curve of a metal in a fairly severely cold-worked state can be approximated by a linear relation

$$\sigma/Y=1+H\epsilon, \quad \dots\dots\dots \quad (19)$$

where Y is the initial yield stress and H is a non-dimensional constant of order unity. We look for a solution giving the various unknown quantities as power series in h/a. We shall take the solution only as far as the second approximation, and it will become apparent that the yield criterion is then required to be satisfied only in first approximation.

Any yield criterion (5), which reduces to von Mises' criterion for stress-states near to a balanced biaxial tension, can then be written as

$$(\sigma_c+\sigma_t)/2\mathrm{Y}\simeq 1+\mathrm{H}\epsilon, \qquad (20)$$

where the equivalent strain ϵ is equal to $\ln(t_0/t)$ to first order.

Elastic strains being neglected, bulging does not begin until the whole sheet is stressed to the yield point. Initially, therefore, $\sigma_c+\sigma_t=2\mathrm{Y}$ everywhere. The only possible stress components satisfying this condition and the second of equations (3) (with $t=t_0$) are

$$\sigma_c=\sigma_t=\mathrm{Y}.$$

This is the first approximation to the stresses. Substitution in the second of equations (4) gives $\theta=pr/2t_0\mathrm{Y}$ to first order. Now

$$\frac{dz}{dr}=-\tan\theta.$$

Hence, after substituting for θ and integrating

$$\frac{z}{h}=\left(1-\frac{r^2}{a^2}\right), \quad p=\frac{4ht_0\mathrm{Y}}{a^2},$$

on introducing the necessary conditions that $z=h$ when $r=0$, and $z=0$ when $r=a$. From (6) and (8) after substituting

$$\theta=\frac{2hr}{a^2},$$

we find that (to first approximation)

$$v=\frac{2hr}{a^2}\left(1-\frac{r^2}{a^2}\right),$$

since $v=0$ when $r=a$. Finally, from (7),

$$\ln\frac{t_0}{t}=\frac{2h^2}{a^2}\left(1-\frac{r^2}{a^2}\right).$$

This completes the solution to first approximation. To this order the bulge is spherical (or rather a paraboloid), irrespective of the rate of work-hardening.

The same sequence of solution is used for the second approximation. The yield criterion (20) becomes

$$\frac{\sigma_c+\sigma_t}{2\mathrm{Y}}=1+\frac{2\mathrm{H}h^2}{a^2}\left(1-\frac{r^2}{a^2}\right).$$

Combining this with the second of (3) we find that the second approximation to the stresses is

$$\left.\begin{aligned}\frac{\sigma_t}{\mathrm{Y}}&=1+\frac{h^2}{a^2}\left[2\mathrm{H}-(\mathrm{H}+1)\frac{r^2}{a^2}\right],\\ \frac{\sigma_c}{\mathrm{Y}}&=1+\frac{h^2}{a^2}\left[2\mathrm{H}-(3\mathrm{H}-1)\frac{r^2}{a^2}\right].\end{aligned}\right\} \qquad (21)$$

We see that $\sigma_t \gtrless \sigma_c$ according as $H \gtrless 1$. Continuing as before, we find successively that

$$\theta = \frac{2hr}{a^2}\left[1+\left\{\left(H-\tfrac{1}{3}\right)\frac{r^2}{a^2}-\tfrac{1}{2}(H+1)\right\}\frac{h^2}{a^2}\right], \qquad (22)$$

$$\frac{z}{h} = \left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{2}(H+1)\frac{r^2h^2}{a^4}\right], \qquad (23)$$

$$\frac{p}{Y} = \frac{4ht_0}{a^2}\left[1+\tfrac{1}{2}(3H-5)\frac{h^2}{a^2}\right], \qquad (24)$$

$$v = \frac{2hr}{a^2}\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{2}\left\{(7H-5)\frac{r^2}{a^2}-(3H-1)\right\}\frac{h^2}{a^2}\right], \qquad (25)$$

$$\frac{u}{a} = \frac{h^2r}{a^3}\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{4}\left\{(7H+1)\frac{r^2}{a^2}-(3H+1)\right\}\frac{h^2}{a^2}\right], \qquad (26)$$

$$\ln\frac{t_0}{t} = \frac{2h^2}{a^2}\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{4}\{(13H-7)\frac{r^2}{a^2}-(3H-1)\}\frac{h^2}{a^2}\right]. \qquad (27)$$

The polar radius of curvature is given by

$$\frac{a}{\rho} = \frac{2h}{a}\left[1-\tfrac{1}{2}(H+1)\frac{h^2}{a^2}\right]. \qquad (28)$$

The equation of a circle passing through the edges and the pole is (to this approximation)

$$\frac{z'}{h} = \left(1-\frac{r^2}{a^2}\right)\left(1+\frac{r^2h^2}{a^4}\right). \qquad (29)$$

On comparing (23) and (29), it is found that the maximum relative difference between the bulge profile and the circle is

$$\text{Max.}\left(\frac{z-z'}{h}\right) = \tfrac{1}{8}(H-1)\frac{h^2}{a^2}, \text{ when } r = \frac{a}{\sqrt{2}}.$$

This vanishes when $H=1$, in agreement with the solution for the special stress–strain curve (12), since $\exp\epsilon \sim 1+\epsilon$ to this approximation. When $H>1$ (which is so for all save exceptional degrees of cold-work) the profile of the bulge lies above the circle (29), and below when $H<1$. Expressed otherwise, the polar circle of curvature cuts the original plane outside the perimeter of the die when $H>1$ and inside when $H<1$. This was found experimentally by Brown and co-workers at all stages up to the pressure maximum ($H \sim \frac{1}{2}$ for the hard copper used by Brown and Sachs and ~ 4 for the hard aluminium alloy used by Brown and Thompson).

From (27) the polar strain is

$$\epsilon = \frac{2h^2}{a^2}\left[1-\tfrac{1}{4}(3H-1)\frac{h^2}{a^2}\right]. \qquad (30)$$

Combining this with (28) we obtain

$$\frac{\rho}{a} = \left[\frac{1}{2\epsilon}+\frac{(H+5)}{16}\right]^{1/2}. \qquad (31)$$

Thus, the effect of the rate of work-hardening on the $(\rho/a, \epsilon)$ relation is comparatively slight up to strains of order 0·3, because of the square root dependence and the small contribution made by the term H/16. (When H=1, equation (31) is the second approximation to the solution of (15), as is to be expected.) The relation between bulge height and polar strain, or between bulge height and polar curvature, is much more dependent on the value of H. We see from (30) that, for a given height, the most severely worked metals (small H) have the largest polar strains but the smallest polar curvatures ; this is in agreement with the observations of Brown and Thompson.

An inspection of equation (28) indicates that a rough criterion for the range of validity of this solution (say to an accuracy of 1 per cent) is

$$(\mathrm{H}+1)\frac{h^2}{a^2} < \tfrac{1}{2}.$$

Thus, when H=1, the solution is probably reliable up to a polar height of $\frac{1}{2}a$. When H=3, a value typical of moderately cold-worked metals, the solution is probably reliable up to a polar height of $a\;2\sqrt{2} \sim 0{\cdot}35a$. It is only when H is less than about $\frac{1}{2}$ that the range of validity probably extends past the pressure maximum which, according to (24), occurs at a polar height of

$$\left[\frac{2}{3(5-3\mathrm{H})}\right]^{1/2} a.$$

Thus, the solution does not appear capable of explaining the secondary bulge observed by Brown and co-workers ; this may indeed be peculiar to annealed metals with a well-rounded stress–strain curve. It has not yet been found possible to obtain any simple approximate solution for such metals.

Acknowledgments.

I wish to thank the Director of the British Iron and Steel Research Association for permission to publish this work, and Mr. I. M. Longman, of the B.I.S.R.A. Solid Mechanics Section for checking the accuracy of the series solution.

References.

Brown, W. F., Jr., and Thompson, F. C., 1949, *Trans. Am. Soc. Mech. Eng.*, **71,** 575.

Brown, W. F., Jr., and Sachs, G., 1948, *Trans. Am. Soc. Mech. Eng.*, **70,** 241.

Gleyzal, A. N., 1948, *Trans. Am. Soc. Mech. Engrs.*, **70,** 288.

Lankford, W. T., and Saibel, E., 1947, *Trans. Am. Inst. Min. Met. Eng., Metals Division*, **171,** 562.

CI. *A Method for the Mapping of Vector Potential Distributions in Axially Symmetrical Systems.*

By G. Liebmann,
Associated Electrical Industries Research Laboratory, Aldermaston, Berks.*

[Received July 27, 1950].

Summary.

The magnetic vector potential distribution in axially symmetrical systems containing current carrying conductors and iron of arbitrary shape can be obtained experimentally with good accuracy using resistor networks. Two networks are possible, one for plotting the flux function rA_θ and one for plotting the vector potential A_θ itself. The latter one is a modification of a field plotting network described previously. Some confirming measurements on a model of an infinitely long coil were carried out.

§1. Introduction.

The determination of the magnetic field due to currents flowing in conductors requires the mapping of the magnetic vector potential. In the presence of iron a mathematical analysis becomes almost impossible, but it was shown by Pérès and Malavard (1938) and by Peierls (1946) that the problem can be solved experimentally with the help of the electrolytic tank analogue in those cases where all currents flow in straight parallel conductors. The problem is then a two-dimensional one, and only one component (*e.g.* the z-component) of the vector potential exists, which satisfies a Poisson equation in x and y. In the electrolytic tank analogue the iron circuits are represented by wax models ; the current conductors, supposed to be perpendicular to the surface of the electrolyte, are represented by a number of probes dipping into the electrolyte, and feeding the appropriate currents into the tank.

This model representation cannot be applied to rotationally symmetrical problems, where all currents flow in circles concentric with the axis of the system. The vector potential has then only an angular component A_θ which has to satisfy this equation :

$$\frac{\partial^2 A_\theta}{\partial z^2}+\frac{\partial^2 A_\theta}{\partial r^2}+\frac{1}{r}\frac{\partial A_\theta}{\partial r}-\frac{1}{r^2}A_\theta=-\frac{4\pi}{10}J, \quad . \quad . \quad . \quad . \quad (1)$$

where J is the current density in the conductors, in amps/cm.^{2}.

Writing $$rA_\theta=\Psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

* Communicated by the Author.

this equation can be put into the form

$$\frac{\partial^2 \Psi}{\partial z^2}+\frac{\partial^2 \Psi}{\partial r^2}-\frac{1}{r}\frac{\partial \Psi}{\partial r}=-\frac{4\pi}{10}r\mathrm{J}, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

It was recently shown by Peierls and Skyrme (1949) that an experimental solution for this function Ψ can be obtained by using two tank measurements, first setting up the model in a tank with equal height of electrolyte, and then in a tilted or wedge tank. The difference between the two measurements can be used as a correction for the results obtained in the first tank, provided the correction is small, *i.e.* measurements are taken not too near the axis.

§2. APPLICATION OF RESISTOR NETWORK METHOD.

A *direct* measurement either of the vector potential A_θ, satisfying equation (1), or of the function $(r\mathrm{A}_\theta)$, satisfying equation (3), can be obtained with a resistor network similar to the type described recently by the present author.

Consider the equation

$$\frac{\partial^2 \mathrm{F}}{\partial x^2}+\frac{\partial^2 \mathrm{F}}{\partial y^2}=\mathrm{G}(\mathrm{F}, x, y). \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

In the earlier paper (Liebmann 1950) it had been shown that the difference equation corresponding to equation (4) is satisfied by a network of interconnected resistances of equal value R_0. The values of the function F are represented by the voltages at the resistor junctions, and the function G is related to the currents I, fed into the junctions from an external source, through the equation

$$\mathrm{G}=-\mathrm{I}\mathrm{R}_0/(\Delta x)^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Now let us apply the transformation

$$\left.\begin{aligned} x&=r^{-k}z \\ y&=\frac{r^{1-k}}{1-k} \end{aligned}\right\} k\neq 1, \qquad \left.\begin{aligned} x&=z/r \\ y&=\log r \end{aligned}\right\} k=1, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where k is any real constant. This transforms the functions $\mathrm{F}(x, y)$ and $\mathrm{G}(\mathrm{F}, x, y)$ into $f(r, z)$ and $g(f, r, z)$ and equation (4) into

$$\frac{\partial^2 f}{\partial z^2}+\frac{\partial^2 f}{\partial r^2}+\frac{k}{r}\frac{\partial f}{\partial r}=r^{-2k}\,g(f, r, z), \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

or with

$$r^{-2k}g=g^* \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

into

$$\frac{\partial^2 f}{\partial z^2}+\frac{\partial^2 f}{\partial r^2}+\frac{k}{r}\frac{\partial f}{\partial r}=g^*. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

If we now choose the four resistances $R_1 \ldots R_4$, joined in the network junction P_0 (see fig. 1) such that

$$\left.\begin{aligned} R_1&=R_3=r^{-k}R_0,\\ R_2&=(r_0^{1-k}-r_2^{1-k})R_0/(1-k),\\ R_4&=r_4^{1-k}-r_0^{1-k})R_0/(1-k), \end{aligned}\right\} \quad . \; . \; . \; . \; . \quad (10)$$

Fig. 1.

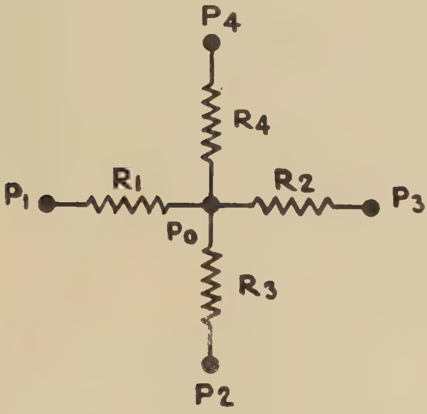
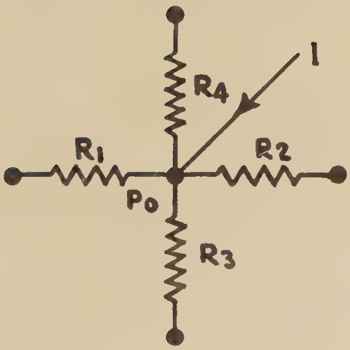


Fig. 2.



R_0 being an apparatus constant, we obtain a network which solves the differential equation (7) to a higher degree of approximation than the difference equation which would replace equation (7) in numerical methods of solution †.

---

† Except for $k \rightarrow 0$, the difference equation as used in most numerical techniques is only correct to second order terms, whereas the specified network is correct to third order terms, whether $k \rightarrow 0$ or not. This point is mostly not very important in regions distant from the axis, but cannot be overlooked as the axis is approached.

In the previous paper (Liebmann 1950) it was shown, through the application of Kirchhoff's law, that if the function F is represented by voltages, then the function G is represented by currents fed into the network points P_0 from an external source (see fig. 2). Similarly, we find that the function g^* is related to the currents I fed in at P_0 by

$$I = -g^*(\Delta z)^2/R_1, \qquad \text{(11}$$

where R_1 is given by equation (10).

§3. RESISTOR NETWORK FOR FINDING (rA_θ).

The two alternative equations (1) and (3) for the vector potential A_θ and for (rA_θ) can therefore be solved by constructing the appropriate networks. Taking first equation (3), we have $k=-1$, and

$$\left.\begin{aligned} R_1 &= R_3 = rR_0, \\ R_2 &= (r_0^2 - r_2^2)R_0/2, \\ R_3 &= (r_4^2 - r_0^2)R_0/2. \end{aligned}\right\} \qquad (12)$$

The currents to be fed in are then

$$I_0 = +\frac{4\pi}{10}\, J(\Delta z)^2/R_0. \qquad (13)$$

Obviously, current would only be fed in at the positions occupied by the conductors. If iron is present, the network must be open-circuited along the boundaries of the iron, the network being correctly terminated there, as discussed in the earlier paper, in order to fulfil the boundary condition $\dfrac{\partial(rA_\theta)}{\partial n}=0$.

§4. RESISTOR NETWORK FOR FINDING A_θ.

It will mostly be more attractive to use a network which solves directly for A_θ. Then $k=+1$, and the previously described network, representing the Laplacian operator for axially symmetrical fields, can be applied. The currents to be fed-in would then have to be $\left(\dfrac{4\pi}{10}J - \dfrac{1}{r^2}A_\theta\right)\Delta z^2/R_1$; this would require the use of the iteration method described before (Liebmann 1949). This need is eliminated by the following modification: Each network junction is connected through an external resistance of value $R_5 = rR_0$ to a point of zero potential (fig. 3).

We then have these values of resistances :

$$\left.\begin{aligned} R_1 &= R_3 = R_0/r_0, \\ R_2 &= R_0 \log(r_0/r_2), \\ R_4 &= R_0 \log(r_4/r_0, \\ R_5 &= r_0 R_0. \end{aligned}\right\} \qquad (14)$$

This network now represents the operator

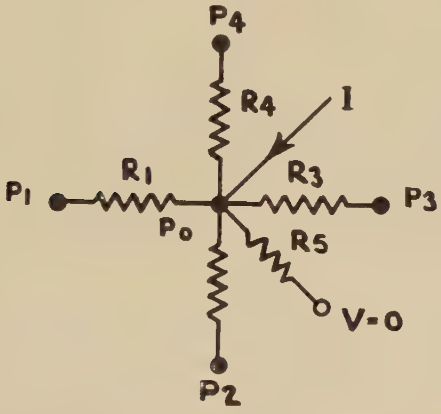
$$\left[\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}\right]$$

which differs from zero only at the position of current carrying conductors. at these positions, the current

$$\mathrm{I}=\frac{4\pi}{10}\,\mathrm{J}\ (\Delta z)^2/\mathrm{R}_1=\frac{4\pi}{10}\mathrm{J}r(\Delta z)^2/\mathrm{R}_0\ . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15)$$

is fed into the network. It is thus possible to measure the vector potential distribution inside current carrying regions within the limitations imposed by the finite size of the interval. Iron circuits are again represented by open circuiting the appropriate boundaries, as $\partial \mathrm{A}_\theta/\partial n=0$, for infinite permeability; the values of the resistances R_5 have to be doubled along the boundaries and quadrupled at interior corners to maintain correct relations at the boundaries. If it is desired to represent correctly a finite value of permeability, the network constants have to be modified, by increasing the value of the constant R_0 in the ratio of $(\mu_{\text{iron}}/\mu_{\text{air}})$, in all resistances within the area occupied by the iron, or by tapping down the voltage in the ratio of $1:\mu$ at the boundary points, using the resistances R_5 as potential dividers, in order to satisfy the boundary condition $\mathrm{A}_{\theta\,\text{iron}}=1/\mu\ \mathrm{A}_{\theta\,\text{air}}$. As previously described, curved boundaries can be simulated by shunting the network resistances; the values of the resistors R_5 have then to be adjusted in a similar way, as mentioned before when referring to the boundaries of iron circuits, taking the change of size of the effective area of the network element $(\Delta z\times\Delta r)$ into account.

Fig. 3.



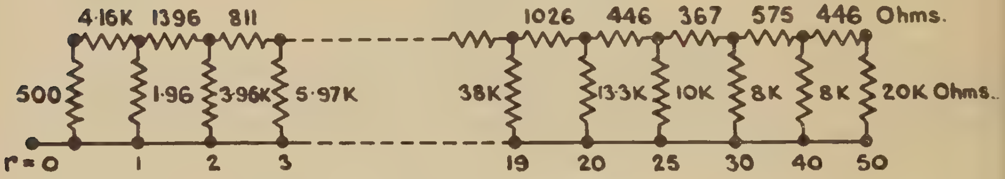
If a problem to be solved involves regions which are open in the direction away from the axis, then it will be necessary to follow the technique applied in similar cases of electrostatic fields. A small scale model is set up on the network, in which the distance of the current carrying conductors from the axis is small compared with the radial extension of the network, and the approximate values of the vector potential at a suitable radial distance r_B are determined. The model is then enlarged so that this relative radial value r_B now corresponds to the edge of the network, and the potential values required at this distance from the axis are forced on to the model by feeding in additional currents at these points. Alternatively, a suitable flux line may be made an equipotential for the voltage. The axis is, of course, an equipotential line of zero potential.

Instead of determining the values of A_θ at the mesh points, one can plot out directly the lines of equal A_θ using the ganged potentiometer arrangement described before (Liebmann 1950).

§5. PRACTICAL REALIZATION AND TESTS OF ACCURACY.

The correctness of the discussed method was confirmed by several experiments with a linear resistance strip. This strip (fig. 4) represented one section of a precision (r, z) resistance network being built in this laboratory. The gradation of the resistance values was on a logarithmic basis (equation (14)), and the values were the same as those employed in the previously described resistance network ($R_0=2000$ ohms, except that the resistance values were kept within about $\pm 0{\cdot}2$ per cent of their nominal values, and that the termination strip was extended to go out from $r=20$, the edge of the network proper, to $r=50$ by a few large steps. Each junction, at distance r from the axis, was connected to the axis ($V=0$) by a resistor of value $R_5=rR_0$, according to equation (14). The finite size of the interval makes the nominal value rR_0 incorrect by a small factor, of the order of 1·8 per cent at $r=1$ and 1·2 per cent at $r=2$, the correction factor diminishing to a negligible amount for greater values of r. Expressing the length r through its equivalent resistance value leads to the corrected value $R_5'=\frac{1}{4}R_0^2\sum_1^4(1/R_i)$; these corrected values were used in the final experiments. The linear resistance strip can be used to investigate problems in which A_θ does not depend on z, *e.g.* the case of infinitely long coils of constant radii. In one such experiment a single layer coil of $r_0=10$ units was assumed, the potential being set to the relative value $A_0=1$ at $r_0=10$. Table I. gives the experimentally found values and the theoretical

Fig. 4.

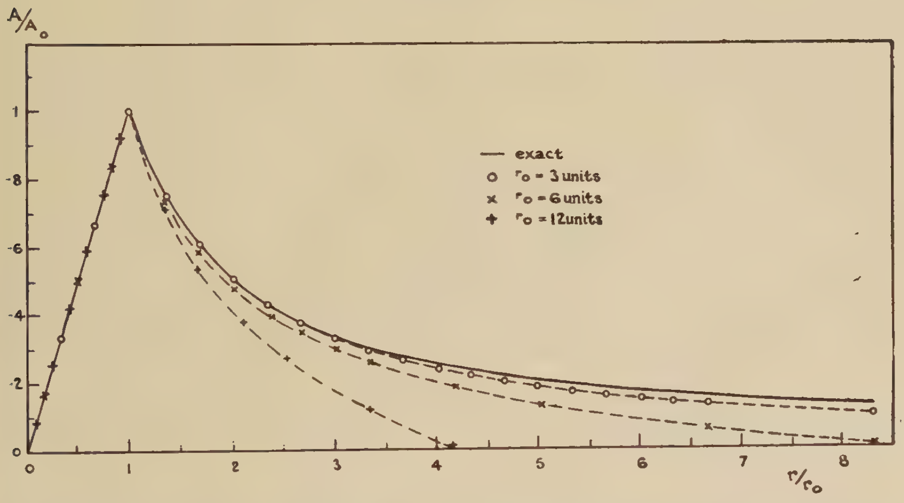


values $A_\theta = A_0 r/r_0$. The agreement between experimental values and theoretical values is very good, particularly in view of the small number of components involved, which does not bring the averaging properties of the network fully into play.

TABLE I.

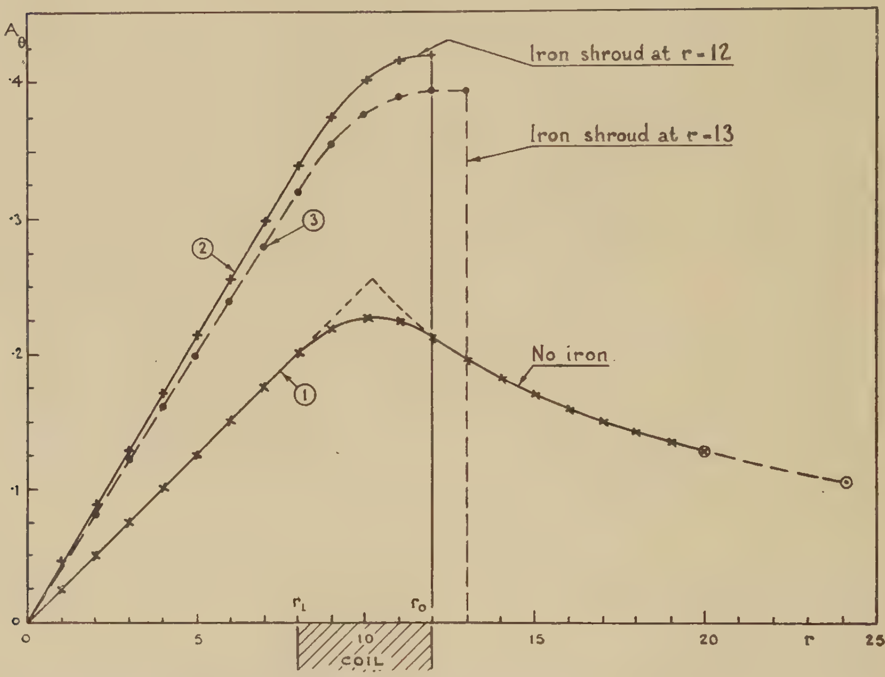
r/r_0	$A_\theta(r)$ exp. Before correction, $R_5 = rR_0$.	$A_\theta(r)$ exp. After correction, $R_5' = \frac{1}{4}R_0^2\sum_1^4(1/R_i)$	$A(r)$ calc.
0	0	0	0
0·1	0·1002	0·1000	0·1000
0·2	0·2012	0·2004	0·2000
0·3	0·3012	0·3004	0·3000
0·4	0·4011	0·4004	0·4000
0·5	0·5009	0·5002	0·5000
0·6	0·6007	0·6002	0·6000
0·7	0·7005	0·7001	0·7000
0·8	0·8003	0·8001	0·8000
0·9	0·9002	0·9001	0·9000
1	1	1	1

Fig. 5.



The vector potential A_θ falls away as $1/r$ outside an infinitely long single layer coil, and an incorrect field distribution will be obtained at some distance from the coil unless the model is made very small in relation to the network, giving low " resolution " near the coil itself. This difficulty can be eliminated by first measuring the field distribution of a small scale model in which the potential $A_\theta = 0$ at $r = 50$ units, and by then setting up the full size model and impressing the potential values found in the first test on the new boundary points. To study the quantitative side of this procedure the field distributions outside single layer coils of radii $r_0=3$, $r_0=6$ and $r_0=12$ mesh units were investigated. The results are plotted in fig. 5, where the full curve represents the theoretical curve $A_\theta(r)=A_0 r_0/r$. For the smallest coil, good agreement between theoretical and experimental values (with $A_\theta=0$ at $r=50$) is found up to about $r/r_0 \simeq 2{\cdot}5$. The error is here $-0{\cdot}004\ A_0$. From graphs of the kind given in fig. 5 information about the appropriate size of the small scale model for the first test can be obtained.

Fig. 6.



The effect of an iron shroud on a coil of inner radius $r_i=8$ cm. and outer radius $r_0=12$ cm. is shown in fig. 6. Curve 1 gives a plot of $A_\theta(r)$ for a coil without iron being present. The boundary at $r=20$ was set to the value obtained from a quarter scale model as described before (circled points). The total current fed into the network was $0{\cdot}2500\ V_0$ ma, where V_0 is the feeding voltage of the network *), or $J=6{\cdot}25\times10^{-5}\times V_0$ amps/cm.^{2}.

Within the coil, the measurements give $A_\theta=0{\cdot}02508\ rV_0$ gilberts, hence $H_z=1/r\ \partial(rA_\theta)/\partial r=0{\cdot}0501_6\ V_0$ oersteds. Calculating H_z direct from the formula $H_z=4\pi/10$ AT/cm. we obtain $0{\cdot}0500\ V_0$ oersteds; this appears to be a quite satisfactory agreement with the experimental results †.

---

* All measurements of voltages and currents were referred to resistance measurements, as this gives a higher accuracy and greater convenience; the feeding voltage V_0 appears then as a common multiplication factor.

† The continuations of the inner and outer branches of the A_θ curve, which obey the (kr) and (k/r) relationships, would intersect at $r=10{\cdot}2$, the centre of gravity of the coil windings.

Placing an iron shroud ($\mu=\infty$) tightly around the coil (at $r=12$) changes the shape of the A_θ curve within the coil considerably (curve 2, fig. 6), and increases the value of A_θ inside the coil to $A_\theta=0{\cdot}04243\,rV_0$, *i.e.* the iron shroud has increased the slope of the A_θ curve, and therefore the magnetic field within the coil, by a factor $1{\cdot}69_2$. If the iron shroud is placed a small distance away from the coil, at $r=13$ (fig. 6, curve 3) the A_θ curve becomes modified, and the factor of increase in field strength in only $1{\cdot}60_0$. A finite permeability ($\mu\simeq1000$) was then simulated, which produced very slight modifications of the A_θ curve near its maximum; the change is too small to be represented in the graph.

§ 6. CONCLUSION.

It is seen from these tests that a satisfactory accuracy can be obtained. As the modifications to existing field plotting networks, satisfying Laplace's equation, can be easily effected, the measurement of vector potential distributions becomes a very practicable proposition, even within current carrying regions and in the presence of iron. The method should therefore have applications not only in the design of betatron and synchrotron magnets, and in similar problems, but also in many ordinary problems of electrical machine design. The same principle can also be applied to the mapping of other vortex fields of rotational symmetry, when the vorticities and positions of the vortex rings are given.

ACKNOWLEDGMENT.

The author is indebted to Dr. T. E. Allibone, F.R.S., for permission to publish this paper.

REFERENCES.

LIEBMANN, G., 1949, *Nature, Lond.*, **164,** 149; 1950 *Brit. Journ. Appl. Phys.*, **1,** 92.
PEIERLS, R. E., 1946, *Nature, Lond.* **158,** 831.
PEIERLS, R. E., and SKYRME, T. H. R., 1949, *Phil. Mag.*, **40,** 249.
PÉRÈS, J., and MALAVARD, L., 1938, *Bull. Soc. Franç. Elect.*, **8,** 715.

CII. *Double Stars due to Cosmic Rays.*

By T. T. Li,
Imperial College, London*.

Summary.

Cosmic ray stars are distributed at random over a photographic plate when distances exceeding 1 cm. are considered. For small distances of the order of 1 mm. the number of close pairs exceeds that expected on a random distribution. The effect is not found on plates exposed at sea-level and there is some evidence that it is more marked if dense matter surrounds the plates. Several mechanisms are suggested to account for the effect but none is free from objection.

§1. Introduction.

Measurements were made (Li and Perkins 1948) on the surface distribution of 1230 stars on 54 in.^{2} of 100μ C2 plates exposed for forty days at 3650 m. altitude. The plates were divided into $(1\times\frac{1}{2})$ in. areas to facilitate searching. It was found that the distribution of star frequencies between these areas conformed to the Poisson distribution, showing that the stars were randomly distributed over areas of these dimensions.

At smaller distances, however, the stars showed a tendency to be in close proximity of each other more frequently than would be expected if they were distributed at random. (See also Leprince-Ringuet and Heidmann 1948.)

Thus, for a random distribution, the number of close pairs of stars, separated by a projected distance in the plane of the emulsion $\leqq r_p$, should be

$$Q(r_p)=\sum_i \frac{Ni(Ni-1)}{2}\cdot\frac{\pi r_p^{2}}{ab},$$

where

Ni =number of stars in the ith area,

a, b=sides of the areas=1 in. and $(\frac{1}{2})$ in. respectively, and r_p is small compared with b.

The actual observed number of close pairs, $P(r_p)$, was compared with the corresponding $Q(r_p)$. It was found that $P(r_p)$ was always greater than $P(r_p)$ up to $r_p=0{\cdot}03$ in., which was then the largest distance measured. Moreover, the ratio $(P-Q)/\sqrt{Q}$ of the observed to the expected deviations was always greater than one. Different mechanisms were suggested to account for this excess of close pairs, which was termed "double stars".

---

* Communicated by Sir George Thomson, F.R.S.

In the present paper, the results are given of further measurements, extended to larger distances r_p and made on plates exposed under various conditions.

§2. EXPERIMENTAL RESULTS.

1. Measurements were made on the following sets of plates :

(*a*) Plates exposed at 3650 m. altitude without any absorbing material immediately above the plates. (A total of 1230 stars on 54 in.^{2}.)

(*b*) Plates exposed at sea-level with no absorbing material immediately above the plates. (140 stars on 27 in.^{2}.)

(*c*) Five boxes of plates exposed in a narrow crevasse ($\sim$10 cm. wide) in a glacier at 3650 m. altitude, at the following depths below the surface of ice : 0 m., 1 m., 2 m., 3·5 m., and 5·5 m. (1126 stars on 77·5 in.^{2}).

(*d*) Four boxes of plates exposed at 3650 m. altitude with the following thicknesses of lead above them : 0 cm., 5 cm., 15 cm., and 35 cm. (374 stars on 16 in.^{2}).

The plates were all 100μ thick of the Ilford C2 type. For the sake of convenience, they will be referred to as the " 3650 ", " sea-level ", " ice ", and " Pb " plates, respectively.

2. Measurements were extended to include all close pairs with a projected separation of $r_p \leqslant (0{\cdot}1)$ in. The formula in (1) for the expected number of close pairs from a random distribution is no longer valid, since r_p is comparable with b now. The following exact formula can be derived:

$$Q(r_p)=\sum_i \frac{Ni(Ni-1)}{2}\cdot\frac{\pi r_p^2}{ab}\cdot\left[1-\frac{4(a+b)r_p}{3\pi ab}+\frac{r_p^2}{2\pi ab}\right]$$

which holds true for $0 \leqq r_p \leqq b$.

In the following table, the values of Q, P, $(\Delta P-\Delta Q)^2/\Delta Q$, and $(P-Q)/(\Sigma Ni)$ are tabulated, using the data of all four sets of plates :

r_p, (1/100) in.	2	3	4	5	6	7	8	9	10
$P(r_p)$	46	107·5	182·5	280	239	505	640	791	936
$Q(r_p)$	37·6	82 ·2	159·1	230·6	329·2	444·1	573·1	717·7	876·5
$(\Delta P-\Delta Q)^2/\Delta Q$	1·9	6·4	0·0	9·5	1·8	0·0	0·3	0·3	1·2
$10^3\,(P-Q)/(\Sigma Ni)$	2·9	8·8	8·2	17·2	21·2	21·2	23·3	25·6	20·8

The last row shows the excess of close pairs per 1000 stars. There seems to be a saturation after $r_p \sim (6/100)$ in. Whether this is real or whether it simply arises from the fact that $Q(r_p)$ increases as $\sim r_p^2$ while $(P-Q)$ increases more slowly with r_p so that the effect gets relatively small at large distances and is lost in the statistical fluctuations, it is not possible to say. (Also $(P-Q)$ must be equal to zero at $r_p=a$, since, by definition, $Q_{r_p=a}=\Sigma Ni(Ni-1)/2=P_{r_p=a}$).

By applying the χ^2-test, it is found that the probability that this excess is only the result of statistical fluctuations is small, $\sim$0·5 per cent. Hence we can conclude that this excess does exist.

3. By comparing results of the different sets of plates, we found that

(*a*) There is no excess of P over Q in the "sea-level" plates, or else the effect is so small that it is outweighed by the statistical fluctuations.

r_p, (1/100) in.	2	3	4	5	6	7	8	9	10
10^3 (P Q)/(ΣNi)	—2·9	7·2	2·1	—5·0	—12·9	6·4	2·1	—8·6	—7·2

It is seen from the above figures that the values of (P$-$Q)/(ΣNi) fluctuate between positive and negative. And the χ^2-test gives a probability of 10 per cent for these values of (P$-$Q) to be due to statistical fluctuations.

(*b*) If we group together (i) all the plates which were exposed at 3650 m. with no absorbing material immediately above them (*i. e.* the "3650" plates, the first box of the "ice" plates, and the first box of the "Pb" plates), and (ii) those which were exposed under a thickness of ice or lead (*i. e.* boxes 2 to 5 of the "ice" plates and boxes 2 to 4 of the "Pb" plates), we notice that the values of (P$-$Q)/(ΣNi) for the latter group are greater than the corresponding ones of the former group :

r_p, (1/100) in.		2	3	4	5	6	7	8	9	10
10^3: (P$-$Q)/	–(ii)	4·4	12·5	14·0	20·0	30·0	24·4	38·6	37·7	38·7
(ΣNi)	–(i)	2·5	6·8	11·0	17·3	19·9	14·9	16·0	21·2	11·8

But the probability that the difference is due to statistical fluctuations is 25 per cent, which is not small enough to enable us to establish the existence of a difference.

§3. Discussions.

1. Possible Errors : Let us examine the possible errors from which an apparent excess (P$-$Q) may arise.

(*a*) *Edge effect.*—The damage of the emulsion near the edge of the plates and at the ruled lines of the areas reduces the effective area of each rectangle to less than half sq. in. by an unknown amount. This might affect the calculation of Q to a significant extent. To eliminate this, let us study only the stars inside an imaginary rectangle of 1 cm. by 2 cm. at the centre of each area. The coordinates of the stars enable us to exclude all stars outside this central rectangle, thus giving a true area of 2 cm.^{2} without any error due to incorrect ruling and damaged emulsion at the edge.

We took, as a random sample, the plates exposed under 0 m., 1 m., and 2 m. of ice. There are 52 areas altogether. The distribution of stars in these 52 1-cm. by 2-cm. rectangles is as follows :

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	17
number of 1×2 cm.^{2} areas with n stars	1	1	0	1	4	6	6	8	8	5	5	2	2	2	1

Thus, considering the central 1 cm. by 2 cm. areas only, we have

$$\Sigma \mathrm{N}i(\mathrm{N}i-1)/2=(1\times 3\times 2+4\times 4\times 3+\ \ldots\ +2\times 13\times 12+1\times 17\times 16)/2$$

$$=1508.$$

$$\therefore\quad \mathrm{Q}(r_p)=1508\,.\frac{\pi r_p^{\,2}}{ab}\left[1-\frac{4(a+b)r_p}{3\pi ab}+\frac{2\pi ab}{r_p^{\,2}}\right].$$

Here $a=2$ cm. and $b=1$ cm. By comparing these expected numbers of close pairs with the observed number inside the central areas, we get the values of $(\mathrm{P}-\mathrm{Q})/(\Sigma\mathrm{N}i)$:

r_p, (1/100) in.	2	3	4	5	6	7	8	9	10
$10^3\,.\,(\mathrm{P}-\mathrm{Q})/(\Sigma\mathrm{N}i)$	8·0	23·0	21·0	15·4	29·8	23·0	11·3	12·8	−14·1

The values of $(\mathrm{P}-\mathrm{Q})/(\Sigma\mathrm{N}i)$ for the same 52 areas but considering the whole $(1\times\frac{1}{2})$ in. area are given below:

4·3	12·8	15·7	16·1	30·2	24·8	28·2	26·8	13·0

A comparison of the two sets of values shows that the "Edge effect" does not contribute much, if any, to the excess (P−Q).

(*b*) *Uneven Thickness of Emulsion.* Uneven thickness of the emulsion has an effect of concentrating stars and thus giving rise to a spurious excess, (P−Q). But it can be shown that even for a variation of thickness of up to 20 per cent, the effect is still very small.

First consider a simple case. Suppose in the area $(1\times\frac{1}{2})$ in., half of the emulsion has a thickness d, and the other half has a thickness d', $d/d'=1{\cdot}2$. Let n and n' be the numbers of stars in the two halves respectively. Then $n/n'=d/d'=1{\cdot}2$.

If the stars were randomly distributed, then the number of close pairs that we should observe in the two halves is

$$\mathrm{P}(r_p)=\frac{n^2}{2}\cdot\frac{\pi r_p^{\,2}}{s}+\frac{n'^2}{2}\cdot\frac{\pi r_p^{\,2}}{s},$$

for small values of r_p, s being $(\frac{1}{2})\times(\frac{1}{2})$ in.^{2}.

On the assumption that the emulsion thickness is uniform, we calculate the expected number of close pairs in the whole area as

$$\mathrm{Q}(r_p)=\frac{(n+n')^2}{2}\cdot\frac{\pi r_p^{\,2}}{2s}.$$

Hence the spurious excess of P over Q due to non-uniform thickness of the emulsion is

$$\mathrm{P}-\mathrm{Q}=\frac{\pi r_p^{\,2}}{4s}\left[2n^2+2n'^2-(n+n')^2\right]=\frac{\pi r_p^{\,2}}{4s}\,.\,(n-n')^2$$

and

$$(\mathrm{P}-\mathrm{Q})/(\Sigma\mathrm{N}i)=(\mathrm{P}-\mathrm{Q})/(n+n')=\pi r_p^{\,2}\,.\,(n-n')^2/4s(n+n').$$

Inserting $n/n'=1\cdot 2$, we get

$$(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)=\frac{\pi r_p^2}{4s}\cdot\frac{0\cdot 04n'}{2\cdot 2}\cdot$$

Assuming a value of $n'=6$, say, we get

r_p, (1/100) in.	2	3	4	etc.
$10^3\,.\,(\mathrm{P}-\mathrm{Q})/\Sigma \mathrm{N}i)$	0·14	0·31	0·56	etc.

These numbers are negligible compared with the observed values of $(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$. If the emulsion consists of two unequal areas of different thicknesses, the spurious excess can be shown in a similar way to be smaller still. Hence we conclude that the contribution due to non-uniform thickness of the emulsion is negligible.

(*c*) *Stars missed owing to Inefficient Searching.*—Inefficiency in searching might also give rise to a spurious excess of P over Q. For instance, if most of the stars missed are those which are at a large distance to any other star, then a spurious (P—Q) can result.

The efficiency of searching depends on the time spent on an area. At the rate (1 in.^{2}/day) with which the plates were searched, we found that about 20 per cent of the stars were missed. The "ice" plates have been searched twice, first at the normal speed and then at half that speed. The overall efficiency thus obtained was 99 per cent.

Assume the extreme case in which all the stars missed were those that were at a distance $>(1/10)$ in. from any other star, so that they would increase the values of $(\mathrm{Q}(r_p)$ but not $\mathrm{P}(r_p)$. Assume a conservative efficiency of searching, in the case of the "ice" plates, of only 98 per cent.

$\bar{n}$=Average number of stars/area=1126/155=7·26.
$\bar{n}'$=Average number of stars/area, corrected for inefficient searching=7·26/0·98=7·41.

$$\bar{n}'(\bar{n}'-1)/\bar{n}(\bar{n}-1)=1\cdot 025.$$

$$\mathrm{Q}'(r_p)=\text{corrected values of }\mathrm{Q}(r_p)=1\cdot 025\ (\mathrm{Q}r_p).$$

We have

r_p. (1/100) in.	2	3	4	5	6	7	8	9	10
$\mathrm{P}(r_p)$	17	43·5	82	127	182·5	229	294	352	416
$\mathrm{Q}'(r_p)$	17·3	38·5	67·5	103·9	147·8	198·2	255·3	318·7	388·1
$10^3\,.\,(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$	—·3	4·3	12·6	20·0	30·1	26·7	33·6	28·9	24·1

We see that after correction, assuming the extreme case, we are still left with large values of $(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$. Moreover, if the inefficiency of searching were the cause of the excess, (P—Q), then there should be a vast difference between the results of the "ice" plates and the other plates—the values of $(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$ for the "ice" plates would be much smaller than those for the rest of the plates. This difference was not found. This shows that the inefficiency in searching does not contribute much to the values of (P—Q).

2. In this section we give some calculations of the possible causes of the "double stars" and compare them with the experimental data.

(*a*) *Regenerative process*—If we assume that the double stars are produced by penetrating radiation, emitted isotropically from the primary stars, producing secondary stars, then the number of double stars with separation $\leqslant r$ is (Li and Perkins 1948).

$$R=(\Sigma Ni)\, q\, \sigma\, L\, [3d/4+d/2 \log_0 (r/d)],$$

for $r>d$, neglecting absorption.

Here d =thickness of the emulsion$=100\mu$,
L =number of nuclei (excluding hydrogen) per c.c. of emulsion $=5{\cdot}6\times10^{22}$,
q =number of penetrating particles from each star,
σ =cross-section of production of secondary stars by the penetrating particles.

Since the values of r we are considering are $\geqslant d$, we can take $r=r_p$. By comparing the calculated values of $R/(\Sigma Ni)$ with the experimental values of $(P-Q)/\Sigma Ni)$, we get

$$q\sigma=(13{\cdot}8\pm5{\cdot}4)\times10^{-24}\ \text{cm}^2.$$

A cross-section of the order of 10^{-24} cm. requires that q, the number of penetrating particles emitted per star, to be ∼14. It seems almost impossible that there are so many particles emitted per star with energy large enough to produce another star.

The values of q for known particles can be estimated as follows: In recent experiments with Ilford G5 emulsion at 3560 m. altitude (Brown *et al.* 1949), 1223 fast particles (with specific ionization $\leqslant 1{\cdot}5$ times the minimum ionization of a singly charged particle) were found to be emitted from 4923 stars. Allowing for an equal number of fast neutral particles, we get a value of 0·5 fast particle per star. Taking the number of grey tracks (ionization between 1·5 to 5 times the maximum ionization of a singly charged particle) per star found for the star group with 7 and 8 heavily ionizing tracks as the average number of grey tracks per star for all stars, and multiplying by 2·25 to allow for neutrons, we get a value of 3 per star. Assuming that half of these have energies large enough to produce another star, we get a total of 2 particles emitted per star, capable of producing a secondary star. Examples of secondary stars produced by fast particles emitted from primary stars have actually been observed with electron sensitive plates. Not all the fast particles end in this way. The contribution they offer, if assumed to be emitted isotropically, can thus account for less than 2/13·8 or 15 per cent of the effect observed.

(*b*) Single penetrating particles producing one or more stars in their passage through the emulsion. In this process, all stars are assumed to be produced by penetrating particles capable of producing more than one star before they lose all or a large part of their energies. A certain directional distribution of the penetrating particles has to be assumed. An examination of a few particular cases will give us some idea of the cross-section required by such a process.

(i) If we assume that the penetrating particles are all coming in a direction perpendicular to the plane of the emulsion, it can easily be shown that the number of double stars is

$$R=(\Sigma Ni)\ d/2\lambda,$$

where λ=mean free path between successive productions of stars in the emulsion.

Taking 20/1000 as the value of $(P-Q)/(\Sigma Ni)$, we get $\lambda=2{\cdot}5$ mm. Or $\sigma=07\times10^{-24}$ cm$\frac{1}{3}$.

(ii) If we assume that the penetrating particles are all coming in a direction parallel to the plane of the emulsion, the number of double stars would be

$$R=(\Sigma Ni)\ r/\lambda,$$

giving a value of $\lambda=9{\cdot}3$ cm. or $\sigma=2\times10^{-24}$ cm^2.

Fig. 1*a*.

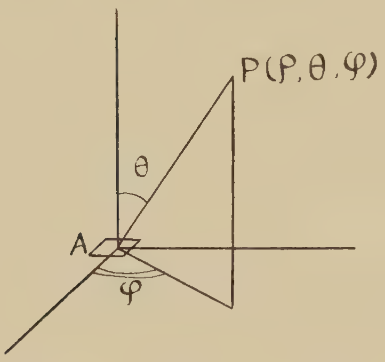
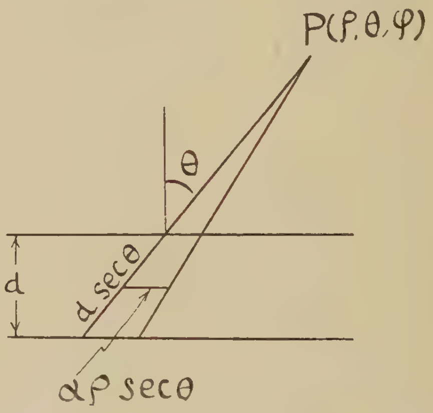


Fig. 1*b*.



(iii) If we assume that the penetrating particles are coming equally probably in all downward directions, then the expression for the number of double stars is simply the same as that for regenerative process except that q is replaced by unity, giving $\sigma=14\times10^{-24}$ cm^2..

Case (iii) is presumably closest to the actual directional distribution, so that with an assumed value of $\sigma=10^{\ 24}$ cm.^{2} this process could account about 7 per cent of the observed double stars.

(*c*) Stars produced by sheafs of particles. If we assume that the stars are produced by sheafs of particles, which are generated in the material surrounding the plates and are emitted isotropically in all downward directions, we can calculate the number of double stars produced.

Let K=number of sheafs generated per unit solid angle per unit volume per unit time,
α =their average conical angle,
S =average number of particles per sheaf,
T =total time of exposure,
λ =m. f. p. for star production in the emulsion,
λ' =m. f. p. for absorption in the surrounding material,
A =area of plates.

Then the number of stars produced in A is

$$(\Sigma \mathrm{N}i)=\int_0^{2\pi}\int_0^{\pi/2}\int_0^{\infty}\mathrm{KT}.\left(\frac{\mathrm{A}\cos\theta}{\rho_2}\right).\,e^{-\varrho/\lambda'}.\,s.\left(\frac{d\sec\theta}{\lambda}\right).\,\rho^2\sin\theta\,d\rho\,d\theta\,d\phi$$

$$=2\pi\,\mathrm{KTA}s\,\lambda'\,d/\lambda.$$

If two members of a sheaf simultaneously produce a star in the emulsion, we have a pair of stars with a projected separation $\leqq \alpha\rho\sec\theta$. Let us take $\alpha\rho\sec\theta \leqq 2\,r_p$ as the condition defining a sheaf so that all stars produced in the emulsion by such a sheaf will have an average projected separation $\leqq r_p$. Then the number of double stars with projected separation $\leqq r_p$ is

$$\mathrm{R}=\int_0^{2\pi}\int_0^{\pi/2}\int_0^{2r_p/\alpha\sec\theta}\mathrm{KT}.\left(\frac{\mathrm{A}\cos\theta}{\rho^2}\right).\,e^{-\varrho/\lambda'}\,\frac{s(s-1)}{2}\left(\frac{d\sec\theta}{\lambda}\right)^2.\,\rho^2\sin\theta\,d\rho\,d\theta\,d\phi$$

$$=2\pi\,\mathrm{KTA}.\,\frac{s(s-1)}{2}\,\frac{\lambda' d^2}{\lambda^2}\int_0^{\pi/2}\left[1-\exp-\left(\frac{2r_p\cos\theta}{\alpha\lambda'}\right)\right]\tan\theta\,d\theta$$

$$\mathrm{R}/(\Sigma\mathrm{N}i)=\frac{(s-1)}{2}.\,\frac{d}{\lambda}.\int_0^{\pi/2}\left[1-\exp-\left(\frac{2r_p\cos\theta}{\alpha\lambda'}\right)\right]\tan\theta\,d\theta$$

$$\equiv\frac{(s-1)}{2}.\,\frac{d}{\lambda}.\,f(r_p)\equiv\mathrm{C}\,.\,f(r_p)$$

The integral can be evaluated graphically if we assign a value for $\alpha\lambda'$. Fig. 2 shows the curves of $\mathrm{C}.f(r_p)$ against r_p for $\alpha\lambda'=(1/2000)$ cm., (1/200) cm., and (1/20) cm. respectively. They are normalized so that $\mathrm{C}.f(r_p)=21/1000$ at $r_p=(1/10)$ in. The factor C is then 0·0027, 0·0040, and 0·0075 respectively. The small circles represent the experimental values of $(\mathrm{P}-\mathrm{Q})/(\Sigma\mathrm{N}i)$. It can be seen that the shape of the curve changes very slowly with the values of $\alpha\lambda'$. It is not a sensitive method to fix the value of $\alpha\lambda'$ by finding the curve which fits the experimental data best.

If we assume $\lambda=18$ cm. (corresponding to a cross-section of 10^{-24} cm.^{2} for the average nucleus in the emulsion), s can be calculated from the value of C. Thus we have

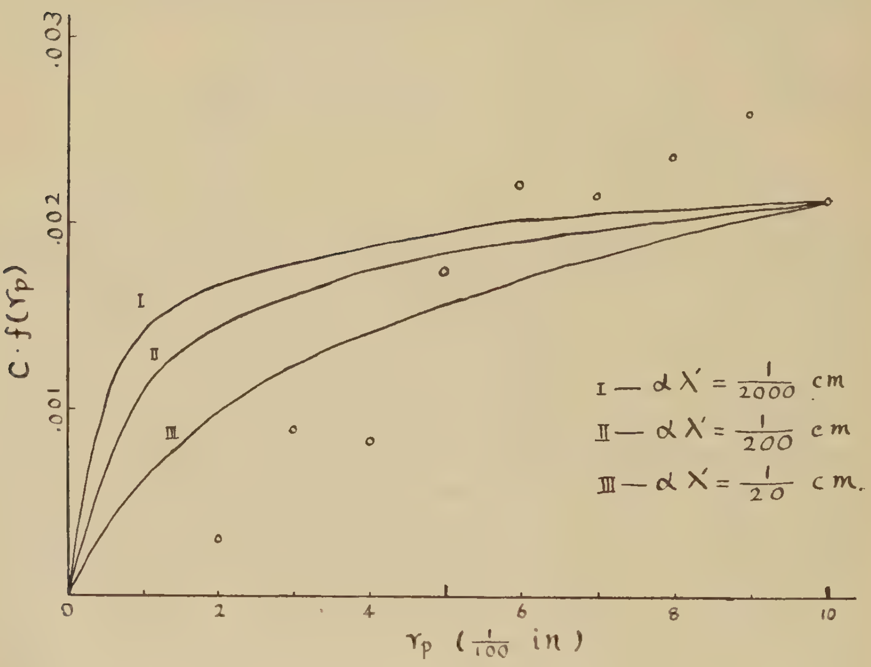
$\alpha\lambda'$	(1/20) cm.	(1/200) cm.	(1/2000) cm.
C	·0075	·0040	·0027
s	28	15	11

We see that if we want to make s reasonably small, $\alpha\lambda'$ would have to be extremely small. It is very difficult, indeed, to explain the double stars by this process even if we assume that all stars are produced in this way. But suppose we accept that $s=15$ and $\lambda=18$ cm. are reasonable, then we can explain $\alpha\lambda'=(1/200)$cm. by assuming a small conical angle, say $\alpha=(1/1000)$ rad., giving $\lambda'=5$ cm. which means that the particles are unstable, having a very short life-time. Thus we have tentatively the following set of values: $\lambda=18$ cm., $\lambda'=5$ cm., $\alpha=(1/1000)$ rad., and $s=15$, which can roughly account for the observed number of double stars.

§4. CONCLUSION.

It was seen in the last section that it is very difficult to account for the observed number of double stars with our present knowledge. The isotropic regenerative process accounts for at most 15 per cent of the effect. Collimated penetrating showers, especially if in cascades, may also account for a fraction of the effect, though it could only be a small one owing to their large conical angles. It is, however, possible that these double stars are due to something hitherto unknown. For instance, neutral particles emitted in large number from primary stars producing secondary stars, or production of stars by sheafs of unstable neutral particles with small conical angles such as appear to produce the close pairs of electrons recently observed (Carlson *et al.* 1950).

Fig. 2.



ACKNOWLEDGMENT.

The author wishes to express his sincere thanks to Prof. Sir George Thomson for his encouragement and help in the completion of this paper.

REFERENCES

BROWN, CAMERINI, FOWLER, HEITLER, KING and POWELL, 1949, *Phil. Mag.*, **40**, 862.
CARLSON, HOOPER and KING, 1950, *Phil. Mag*, **41**, 701.
LEPRINCE-RINGUET and HEIDMANN, 1948, *Nature, Lond.*, **161**, 844.
LI and PERKINS, 1948, *Nature, Lond.*, **161**, 844.

CIII. *Communication Theory and Physics.*

By D. Gabor,
Imperial College, London*.

[Received August 10, 1950.]

Summary.

The electromagnetic signals used in communication are subject to the general laws of radiation. One obtains a complete representation of a signal by dividing the time-frequency plane into cells of unit area and associating with every cell a " ladder " of distinguishable steps in signal intensity. The steps are determined by Einstein's law of energy fluctuation, involving both waves and photons.

This representation, however, gives only one datum per cell, *viz.*, the energy, while in the classical description one has two data : an amplitude and a phase. It is shown in the second part of the paper that both descriptions are practically equivalent in the long-wave region, or for strong signals, as they contain approximately the same number of independent, distinguishable data, but the classical description is always a little less complete than the quantum description. In the best possible experimental analysis by an electronic device the number of distinguishable steps in the measurement of amplitude and phase is only the fourth root of the number of photons. Thus it takes a hundred million photons per cell in order to define amplitude and phase to one per cent each.

Communication theory has up to now developed mainly on mathematical lines, taking for granted the physical significance of the quantities which figure in its formalism. But communication is the transmission of physical effects from one system to another, hence communication theory should be considered as a branch of physics. Thus, it is necessary to embody in its foundations such fundamental physical data as the quantum of action, and the discreteness of electric charges. This is not only of theoretical interest. With the progress of electrical communications towards higher and higher frequencies, we are approaching a region in which quantum effects become all-important. Nor must one forget that vision, one of the most important paths of communication, is based essentially on quantum phenomena.

Some years ago, I have proposed a mathematical framework for the representation of signals (Gabor 1946). I have been rightly criticized for having left out noise, which is an essential feature of all communications. This will be remedied here, and at the same time the description will be brought in line with modern physics. But as the mathematical frame will serve as a useful foundation, it will be necessary to give first a short review of it.

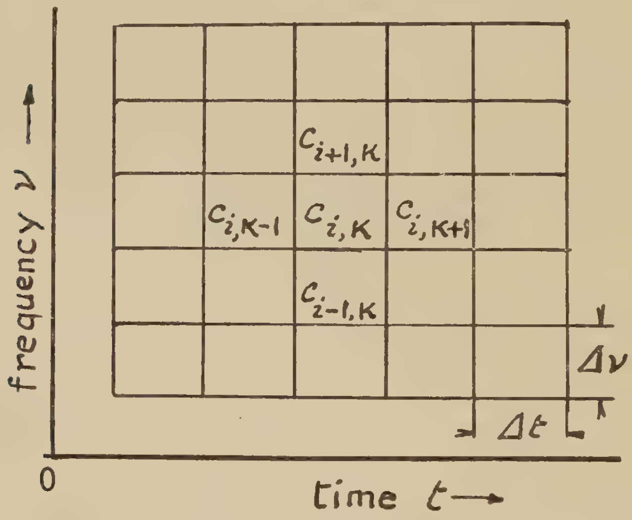
---

* Communicated by the Author.

§1. Classical Representation of Signals.

The previous work started from the observation that the description of a signal in the conventional way, as a continuous function of time, is redundant and non-physical. A continuous function contains in any interval, however small, an infinity of data, corresponding to an infinite range of frequencies. A similar objection can be raised against the Fourier representation, which involves infinite time. In the new description one considers the signal simultaneously as a function of frequency *and* time. It is convenient to use only positive frequencies in the description. This can be done by introducing a certain complex function, whose real part is the physical signal. (The theory of these complex or " analytical " signals has in the meantime received interesting additions by the work of J. A. Ville (1948)). If the time-frequency half-plane, fig. 1, is divided, by any network, into cells of unit area, $\Delta t \Delta \nu = 1$, one finds that the signal in any domain containing a sufficiently large number of cells is fully described by associating two real data, or one complex datum, with every cell. In other words, each cell has two degrees of freedom.

Fig. 1.

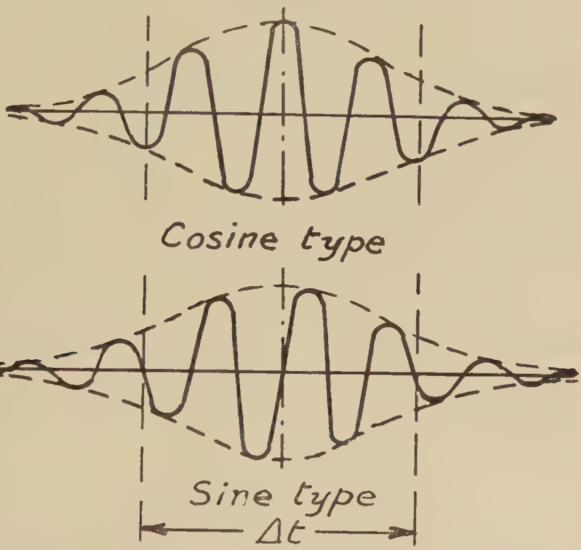


Information diagram. The time–frequency half-plane is divided up into cells of unit area and an elementary signal is associated with each, with a coefficient c_{ik}.

One can represent an arbitrary signal in an infinity of ways as a linear function of certain " elementary " signals, associated with the individual cells. There is, however, one description of particular interest, in which the elementary signals are harmonic functions, modulated with a " gaussian " signal, *i.e.*, they have envelopes of probability shape.

(Fig. 2.) These share with other functions the property that their Fourier transforms have the same shape, but they are unique in that respect that the product of their "effective" duration and of their effective frequency width is the smallest possible for any function. Thus, it can be said that these elementary functions overlap as little as possible. They have also the advantage that the familiar concepts "amplitude" and "phase" can be used in connection with them in the same way as with infinite harmonic functions. The complex elementary function is Cos + $_j$Sin, if we denote by Cos the even, and by Sin the odd type of real elementary signal. Multiplying these with suitable complex coefficients c_{ik}, as indicated in fig. 1, we can represent any arbitrary signal. Time-description and frequency-description (Fourier integral)

Fig. 2.



Elementary signals of the "cosine type" (even), and of the "sine type" (odd).

can be considered as extreme special cases of this representation. In the first case the even elementary signal degenerates to a delta-function, and the odd one to its derivative, in the second case both become ordinary harmonic functions.

The "matrix" representation, illustrated in fig. 1, is proof against the objections raised against the pure "time" or "frequency" descriptions, but it does not go far enough. The infinity of data has been reduced to a finite number, but these data, *i.e.*, the coefficients c_{ik}, are still supposed to be *exactly* defined. But a single exact datum still contains an infinite amount of information, *i.e.* an infinite number of "yeses or noes". In reality, of course, these amplitudes, like every physical datum, have a certain amount of uncertainty or "noise." This has been taken into consideration in the mathematical theories of Shannon (1948) and Tuller (1949), where the noise amplitudes are assumed as known. But we cannot be satisfied with this in a physical theory. Even if all accidental imperfections of the instruments are eliminated, there remain certain basic uncertainties, which we are now going to investigate.

§2. STATISTICAL PROPERTIES OF THE INFORMATION CELL.

In order to connect the mathematical scheme with physical reality, we must first observe that every physical signal has a certain energy associated with it. We can also associate a certain energy with every cell*, and, for simplicity, we will say that it is " contained " in the cell.

We will limit the discussion to electric communications, though most of our results will apply equally well to sound, or, in short, to communication by any quantity which is considered as continuous in classical physics. We observe first that all electric signals are conveyed by *radiation*. Even if lines or cables are used in the transmission, by the Maxwell-Poynting theory, the energy can be located in empty space. Hence we can apply to our problem the well-known results of the theory of radiation.

For simplicity, we consider our communication system as having the uniform temperature T. The uncertainty connected with the concept of temperature will produce certain fluctuations in the energy of the cells, which we can calculate by the rules of statistical thermodynamics once we know the law for the mean thermal energy $\bar{\epsilon}_{\mathrm{T}}$ of a cell, in function of the temperature. This we obtain at once, if we observe that every cell in the signal has two degrees of freedom, of which only one *counts* for the purpose of statistics, as the other is of the nature of a phase. Thus, by Planck's law

$$\bar{\epsilon}_{\mathrm{T}}=\frac{h\nu}{\exp(h\nu/k\mathrm{T})-1}, \quad \ldots\ldots\ldots \quad (1)$$

where ν is the mean frequency of the cell. In other words, we identify every information cell with a " Planck oscillator ".

This requires perhaps a little more explanation, as physicists are less familiar with a discussion of radiation in terms of frequency and time than in terms of frequency and space. But the first case is immediately reduced to the second if we imagine the signal propagating with the velocity c, and plot the information diagram against the length ct instead of against the time t. The state of the field in such a linear system

---

* The energies of the elementary signals of the type discussed will not add up exactly to the energy of the whole signal, because they are not quite orthogonal, but the error will vanish in the mean over large number of cells. One can, however, to meet objections, consider instead an orthogonal set of elementary signals, such as the signals with " limited spectrum ", introduced by Shannon (1949) and by Oswald (1949, 1950) which have uniform spectral density inside and zero outside a frequency strip. This makes no difference to the following discussion, as no reference will be made to any special type of elementary signal.

(in which we consider one state of polarization only) can be represented by two systems of progressive waves in opposite directions, only one of which represents the signal in which we are interested. We count the degrees of freedom in this wave system—which is what we have done in the last section—and give it the energy $\bar{\epsilon}_T$ for each amplitude, disregarding the phases. This is the application to the linear case of von Laue's well-known derivation of Planck's law for the radiation density by superposition of plane waves (von Laue 1914).

The thermal energy $\bar{\epsilon}_T$ does not in itself represent "noise", *i.e.*, an uncertain disturbing factor. It becomes disturbing only by its fluctuations. In order to obtain the mean square fluctuations of the energy, we apply Einstein's formula

$$\overline{\delta\epsilon_T^2} = \overline{(\epsilon_T - \bar{\epsilon}_T)^2} = kT^2 \frac{d\bar{\epsilon}}{dT} T, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

which gives

$$\overline{\delta\epsilon_T^2} = h\nu\bar{\epsilon}_T^2 + \bar{\epsilon}_T. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Einstein's interpretation of this equation is well known (*cf.* Born 1949). The second term has been identified by H. A. Lorentz with the fluctuations due to the interference of waves with random phases.* But the first term suggests that the energy is concentrated in light quanta or photons of energy $h\nu$, which fluctuate in any element as if they were independent particles. As both effects are present simultaneously, and are acting as if they were independent of one another, it is not possible to make a simple physical picture of the process. But fortunately this is not necessary. Applying Einstein's argument to the case when a signal is present, so that the mean energy $\bar{\epsilon}$ in the cell exceeds the thermal mean energy $\bar{\epsilon}_T$, we obtain in Appendix I.

$$\overline{\delta\epsilon^2} = \overline{(\epsilon - \bar{\epsilon})^2} = h\nu\bar{\epsilon} + 2\bar{\epsilon}\bar{\epsilon}_T - \bar{\epsilon}_T. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Expressing the energy by the number of photons N in the cell, so that $\epsilon = Nh\nu$, $\bar{\epsilon}_T = \bar{N}_T h\nu$, this becomes

$$\overline{\delta N^2} = \overline{(N - \bar{N})^2} = \bar{N}(1 + 2\bar{N}_T) - \bar{N}_T^2 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

with

$$\bar{N}_T = 1/[\exp(h\nu/kT) - 1]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

---

* In technical theories of thermal noise, it is usually forgotten that it is not the noise power but its fluctuations which cause the disturbance. But if the quantum effect is small and the second term in equation (3) predominates, the r.m.s. value of the noise power fluctuation is equal to the noise power itself, hence this error is without consequences.

From equation (3), neglecting the quantum term, one can easily derive Nyquists well-known rule (Nyquist 1928) that a resistance R can be considered as containing a "noise generator" with a mean square electromotive force

$$\overline{E^2} = 4kTR\Delta\nu.$$

The proof can be given in the same form as Nyquist has done, by substituting a cable with wave impedance R for the resistance. But one must add the condition that not only the noise power, but also its fluctuations follow the same rule in the resistance as in the cable. This is no arbitrary rule; the necessity of a universal law of mean square fluctuations follows directly from the second principle, as Szilárd (1925) has shown.

$\overline{N}_T$, the number of " thermal " photons per cell, is a large number for the frequencies used in electrical communications. One can use the approximate formula

$$\overline{N}_T=\frac{kT}{h\nu}=\frac{kT}{hc}\lambda=0\cdot7\lambda T$$

which gives $\overline{N}_T=210$ for a wavelength of 1 cm. and a temperature of 300° K. On the other hand for visible light, $\lambda=5\cdot10^{-5}$ cm., the approximate formula

$$\overline{N}_T=\exp(-h\nu/kT)=\exp(-1/0\cdot7\lambda T)$$

gives $\overline{N}_T \doteq e^{-94} \doteq 10^{-39}$. Hence for visible light, at ordinary temperatures, there is practically no thermal noise, and the fluctuation becomes pure " quantum noise ", which follows the law $\overline{\delta N^2}=\overline{N}$.

These results enable us to construct a complete physical representation of a signal. We see that the state of an information cell is completely determined by a stochastic number N, the number of photons. We can now construct a scale or "*ladder*" *of distinguishable states*, on which every step corresponds to a reasonably ascertainable difference. It is an evident suggestion to adopt the r.m.s. fluctuation of N as the unit step.* With this convention, the number of steps distinguishable below a maximum level N_m is, by equation (5),

$$S=\int_{\overline{N}_T}^{N_m}\frac{dN}{(\overline{\delta N^2})^{1/2}}=\int_{\overline{N}_T}^{N_m}\frac{dN}{[N(1+2\overline{N}_T)-\overline{N}_T^2]^{1/2}}$$

$$=\frac{2}{1+2\overline{N}_T}\{[N_m(1+2\overline{N}_T)-\overline{N}_T^2]^{1/2}-[\overline{N}_T(1+\overline{N}_T)]^{1/2}\}\approx\frac{2N_m^{1/2}}{(1+2\overline{N}_T)^{1/2}}. \quad . \quad (7)$$

The last formula is valid for large signals.

In the useful terminology introduced by D. M. MacKay (1950), S is the " proper scale " of the photons. Fig. 3 illustrates the representation of a signal in three dimensions, with the photon scale at right angles to the time–frequency plane. If instead we plotted $\log_2 S$, the ordinates would give directly the equivalent number of binary selections or " bits ". This is the number of "yeses" or "noes" required to fix the position of the signal on the ladder.

It is of interest to enquire about the minimum energy required for the transmission of the first "bit" of information. By our convention, this cannot be less than the r.m.s. value of the thermal energy fluctuations, though it can be more. Combining equations (1) and (3), one obtains

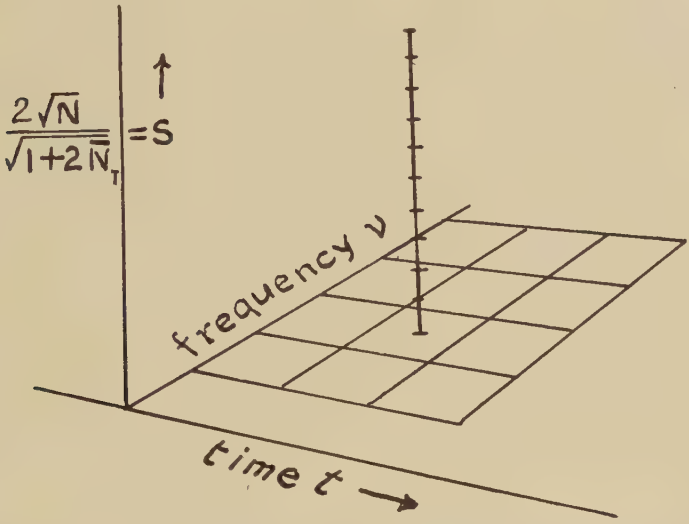
$$\epsilon_{\min}=(\overline{\delta\epsilon_T^2})^{1/2}=\frac{h\nu}{2\sinh(h\nu/2kT)}. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

---

* By the theorem of Bienaymé and Tchebycheff, the probability of an error k times exceeding the r.m.s. error is smaller than $1/k^2$, whatever the law of the fluctuations may be.

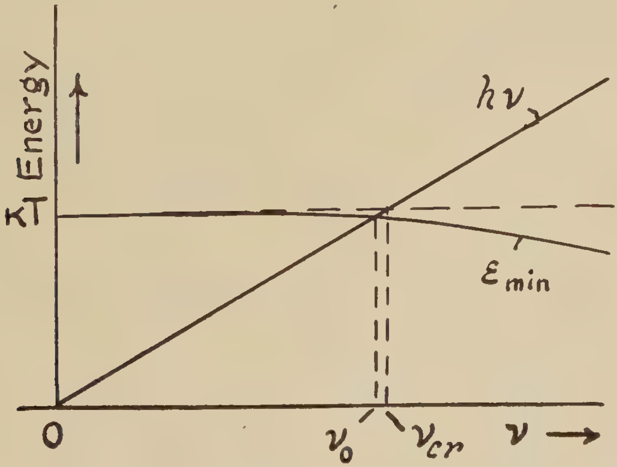
But, on the other hand, this energy cannot be less than one quantum, $h\nu$. Thus the energy required for the first " bit " is either $\epsilon_{\min}$ as given by equation (8), or $h\nu$, whichever is the larger of the two. As shown in fig. 4, the two lines cross at $h\nu_0 = 0{\cdot}96kT$, which is only slightly less than kT.

Fig. 3.



Representation of an arbitrary signal. A ladder or " proper scale " of distinguishable values is erected in every cell, on which the occupation is marked off.

Fig. 4.



The energy required for the transmission of the first " bit " of information.

Thus up to a critical frequency ν_{cr}, an energy kT is sufficient for the first step, but no communication is possible with an energy of less than kT.

The interesting feature of this result is its generality; it applies to the unknown processes in the nervous system as well as to electrical communications.

§3. STATISTICS OF SIGNALS.

Up to this point, we have directed our attention only to one cell. A short digression on signals and ensembles of signals may not be out of place before we return to the physical analysis of our results.

A signal is a system of, say, n cells, preferably, but not necessarily, contiguous in the information plane. We consider a large number of such systems in a stationary transmission, in which they differ only by their position in time, not in frequency, and we speak of them as of an "ensemble". Evidently the analogy with statistical mechanics is not very perfect. In statistical mechanics, we can either follow a system in an ensemble over a long time, or we can look at all systems in the ensemble simultaneously, while here we can take only the first view. It is also somewhat questionable whether the often used expression "ergodic" is justified. In its original sense, due to Willard Gibbs, it means that each system (each signal) spends equal times in all states compatible with a given energy, which is not true for most stationary transmissions usually considered as "ergodic" in communication theory. Thus, we prefer to avoid this term.

The most important mean value in such an ensemble is that of the *entropy*. In order to clarify the connexions between physics and communication theory, it may be useful to consider this problem in two stages. In the first, we consider all configurations of the system of n cells, compatible with the energetic conditions of the transmission as equally probable, and define the entropy as $k \log \mathrm{P}$, where P is the number of all these possible configurations. In the second stage, however, we give them different probabilities or "weights". The first stage is in close connection with physics; actually it is the problem of calculating the entropy of an "ergodic" system, in the original meaning of the word. According to quantum statistics, all simple, accessible states have equal probability, and the levels of Planck oscillators are, of course, simple. (*Cf.* Jordan 1933.)

As an example, let us estimate the mean entropy of groups of n cells in a transmission in which the mean energy level is $\overline{\mathrm{S}^2}$, and the r.m.s. deviation from the mean is $\varDelta \mathrm{S}_n$, for brevity. (The suffix n had to be added, as the standard deviation is dependent on the size of the group.) For simplicity, we assume $(\varDelta \mathrm{S}_n)^2 \ll \overline{\mathrm{S}^2}$, *i.e.*, a small degree of "modulation". This allows us to equate the number P of possible states to the number of points with positive, integer coordinates inside a spherical shell in n-dimensions, whose mean radius R is given by

$$\mathrm{R}^2 = n\overline{\mathrm{S}^2},$$

while its thickness is

$$2\varDelta \mathrm{R} = n\varDelta \mathrm{S}_n/\mathrm{R}.$$

Well-known formulæ for the volume of a n-dimensional sphere give

$$\mathrm{P}=\frac{(\pi/4)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n+1)}\,n\mathrm{R}^{n-1}2\Delta\mathrm{R}$$

$$=\frac{(\pi/4)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n+1)}\,n(n\overline{\mathrm{S}^2})^{1/2n}\,\frac{\Delta\mathrm{S}_n}{\overline{\mathrm{S}^2}}\approx\frac{1}{\sqrt{\pi}}\,(\tfrac{1}{2}e\pi\overline{\mathrm{S}^2})^{\frac{1}{2}n}\,\frac{\sqrt{(n)}\,\Delta\mathrm{S}_n}{\overline{\mathrm{S}^2}}. \quad . \quad (9)$$

We have used Stirling's formula to approximate the Γ-function for large arguments. For sufficiently large n, however complicated the signals, they must behave as if they were independent, and we must have asymptotically

$$n\,.\,(\Delta\mathrm{S}_n)^2\rightarrow\text{const.};$$

hence, apart from a constant, the entropy *per cell* goes to

$$k\log\,(\tfrac{1}{2}e\pi\overline{\mathrm{S}^2})^{1/2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

a result also obtained by Shannon (1949).

Apart from the factor k, the entropy (10) is also a measure of the information capacity per cell of the transmission, in accordance with its definition as the logarithm of the number of possible, equally probable, selections. Thus, the concepts of information capacity and of entropy are closely parallel, even identical.

Consider now for a moment the central problem of statistical communication theory, which has been investigated by Shannon, the communication of symbols, as supplied by some source. If the source supplied them with equal probability, one could connect the possible and equally possible states of the signal with the symbols by an infinity of possible codes, and transmit information at the maximum rate, given by (10) multiplied by the number of cells per second. But if, as it is mostly the case, the symbols have different probabilities p_i, it would be uneconomical to give them equal shares in the available signal entropy. Those which occur more often, and hence carry less information, should be given less than the more unexpected ones. In this way, though it is not, of course, possible to make the channel carry more information than given by (10), it is possible to make it carry *more symbols*, each symbol in the average being worth less. Shannon (1948) has shown that in the case of optimum coding each bit of channel capacity can carry 1/H symbols, where H is defined as

$$\mathrm{H}=-\Sigma p_i\log p_i\ \text{bits/symbol}, \quad . \quad . \quad . \quad . \quad . \quad (11)$$

i.e., as the mean value of a symbol in terms of binary selections. This "*selective entropy*" is defined by the source of symbols, and it is at an appreciable remove from the physical entropy of the signal previously discussed, as may be already seen from the fact that it appears as its divisor. The interesting properties of the expression (11), demonstrated by Shannon, follow rather from its mathematical form than from its intrinsic relation with Boltzmann's "H".

There exists, however, a general limiting relation between selective information and entropy in the thermodynamical entropy of the source considered as a physical system, which was revealed by Szilárd (1929). This is that a signal corresponding in information value to an s-fold selection enables the receiving system to reduce the entropy of the transmitter by a *maximum* of $k \log s$. It follows, as he proved in detail, that any mechanism acquiring this information must increase the entropy by a *minimum* of $k \log s$, thus safeguarding the second principle of thermodynamics.

§ 4. THE CONNECTION BETWEEN QUANTUM REPRESENTATION AND CLASSICAL DESCRIPTION.

Comparing the representation of a signal, illustrated in fig. 3, with the classical, mathematical description, it appears that we have lost something. Previously we had two data per cell; we are left now with one only, a function of the quantum number N, which corresponds to the energy. What has happened to the phase ?

It is not difficult to give an answer to this question on general lines. N is not an exact datum but a stochastic number, which represents a finite amount of information. But between one datum of finite accuracy and two, there is no transcendental abyss of the kind which exists between a simple and a twofold infinity. One finite datum can very well contain two independent, finite data, provided that their aggregate information is not more than the original. It will be shown that this is indeed the case, and that the maximum amount of information on amplitude and phase are both contained in the single photon scale.

An electromagnetic signal can be physically analysed in various ways, but it will be instructive to consider first extreme cases only. One extreme is a *counter*, an instrument which records single photons. It follows immediately from Heisenberg's uncertainty principle that the "time resolution" of such an instrument cannot be better than a whole cycle, hence the phase remains entirely unobservable. The other extreme is any classical field-measuring instrument, capable of recording the electromagnetic field in the signal as a function of time. It may be called for brevity a proportional *amplifier*, as amplification of weak signals is one of its essential functions. There is no need to consider intermediate instruments, as it will be seen that in a certain range of very weak signals every classical amplifier operates as a sort of "proportional counter".

One limit for the operation of any such instrument is given by the well-known uncertainty relation of the quantum theory of radiation (*cf.* Heitler 1944, p. 68)

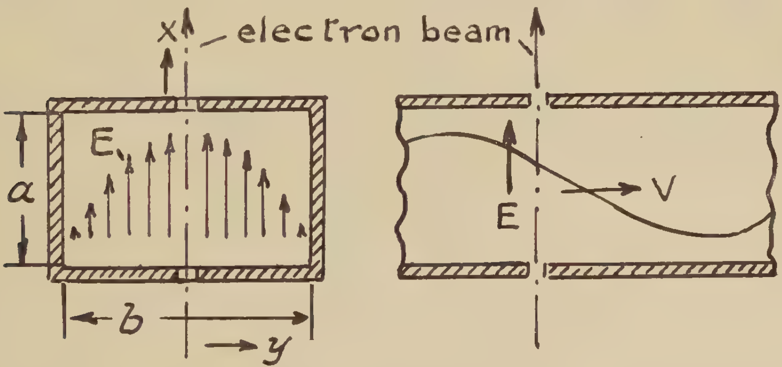
$$\Delta N \Delta \phi \geqslant 1, \quad \ldots \ldots \ldots \ldots \quad (12)$$

where ΔN is the uncertainty in the photon number N, and $\Delta \phi$ the uncertainty in the phase ϕ. But this gives merely an upper limit in our case, as we want to determine amplitude and phase *simultaneously*, each as accurately as possible. It will be seen that, in fact, the limiting accuracy which is practically obtainable is much below what might be expected from the inequality (12).

We will approach the problem by a detailed analysis of a particular type of proportional detector-amplifier. Subsequently, we will try to improve its performance to the extreme limit. It will then be found that all special features of the device vanish from the formulæ, thus the final result can be considered as of general validity.

Assume that the electromagnetic signal is introduced into a rectangular wave guide, in the TE_{10} mode, *i.e.*, with an electric field in the x-direction, independent of the x-coordinate, which is at right angles to the direction of propagation z. An electron beam passes through two small holes along the x-axis, in the plane of maximum intensity, with velocity v and current intensity J. The alternating acceleration and retardation of the electrons produces velocity modulation in the beam, which by well-known methods can be translated into current modulation, *i.e.*, producing an alternating component superimposed on the mean value J. One method utilizes the " bunching " of the electrons which takes place at a certain distance from the guide, but one can also deflect the beam by a constant field and make it play between two collecting electrodes close together. In either case, one obtains an alternating current which in first approximation is proportional to the d.c. current and to the relative accelerations and retardations suffered by the electrons in the wave guide.

Fig. 5.



Analysis of an electromagnetic signal in a wave guide by an electron beam.

Let us measure the energy exchange between the electrons and the field in quantum units, $h\nu$, where ν is the mean frequency of the signal. (It will soon be seen that, in order to make the exchange intense, the waveband of the signal must be made so narrow that it is permissible to take the arithmetic mean.) Let $\overline{\mathrm{N}}$ be the mean number of photons in an information cell, which passes through a cross-section of a guide in a time $\Delta t = 1/\nu$. During this time $\bar{N} = J\Delta t/e$ electrons pass through it, and exchange in the mean $\bar{n}$ quanta with the field, either by losing or by gaining energy. The positive number $\bar{n}$ is the essential parameter of the process. If $\bar{n}$ is a large number, which is possible only if $\bar{N}$ is also large, the interchange will be essentially classic; if $\bar{n}$ is small, quantum phenomena will dominate.*

The detailed calculations may be found in the Appendices II. and III., only the results will be discussed here. The first contains the classical theory of exchange, valid for large $\bar{n}$ and $\bar{N}$, the second a wave-mechanical calculation valid for small exchange parameters. In the classical theory the result is

$$\bar{n}^2 = \frac{32}{\pi^3}\left(\frac{2\pi e^2}{hc}\right)\frac{\sin^2\theta}{\theta}\,\frac{v}{\nu b}\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{-1/2}\frac{\Delta\nu}{\nu}\,\bar{N}. \quad . \quad . \quad (13)$$

Thus the exchange parameter is proportional to the square root of the photon number, as may be expected. Of the dimensions of the wave guide the width b appears explicitly in the factor of $\bar{N}$, while the depth a is contained in the angle

$$\theta = \pi\nu a/v \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

which is one-half of the " transit angle ". c is the velocity of light. In addition, there appears a factor which is the reciprocal of

$$\frac{hc}{2\pi e^2} = 137,$$

the fundamental number which connects photons with electrons.

Anticipating that the best measurement will require an intense interchange, we now try to increase the coefficient of $\bar{N}$ in equation (13) by all available means. First we make the factor $\sin^2\theta/\theta$ a maximum. This is 0·723 and is obtained with $\theta = 67°$, *i.e.*, a transit angle of 134°. This disposes of the depth a of the wave guide. The optimum width b is determined by the condition that the group velocity

$$U = c\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}$$

must be as small as possible. But the smallest value is reached when U vanishes at the low-frequency limit, $\nu - \frac{1}{2}\Delta\nu$, of the band. Substituting those values into equation (14), one obtains in the optimum case

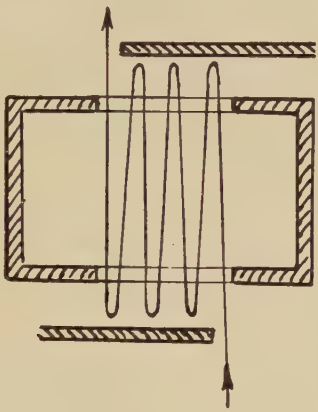
$$\bar{n}^2 = \frac{1\cdot5}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}}\bar{N}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

---

* Quantal energy exchange between electrons and the field in a wave guide at high quantum numbers has been previously discussed by Lloyd P. Smith (1946), but we cannot agree with most of his results. Monokinetic electrons and exchange of sharply defined quanta on the one hand, well-defined entrance phases and short transit times on the other, are mutually exclusive phenomena by the Uncertainty Principle; hence we believe that only certain averages over Smith's detailed results have physical significance.

All special features of the device have vanished in this formula, apart, perhaps, from the unimportant factor, 1·5. But it is evident that the factor of $\overline{\mathrm{N}}$ must always be much smaller than unity, while its best value as will be shown later, is just unity. There exists, however, a further possibility for improving the performance. Assume that we can make each electron perform repeated passages through the guide, each transit, in the opposite direction exactly half a cycle after the last. (This is possible in principle, as the optimum transit angle is about 134°.) If the frequency were known beforehand, the number of passages would be limited only by the consideration that, by repeated gains or losses, the electron would be bound to fall out of synchronism. But if the signal had a single frequency, known in advance, there would of course be no communication. However, even if the frequency is known beforehand only within $\nu \pm \frac{1}{2}\Delta\nu$, one can make the number P of passages as great as $\nu/\Delta\nu$, without risking a phase error of more than $\pm\frac{1}{2}\pi$, and it can be shown that these passages are almost of equal value, so that $\bar{n}$ is increased very nearly by a factor P. The number of passages required to make the coefficient of $\overline{\mathrm{N}}$ in equation (15) unity is

Fig. 6.



Repeated passages.

$$\mathrm{P}_{\mathrm{opt}} = 9{\cdot}5\left(\frac{v}{c}\right)^{1/2}\left(\frac{\nu}{\Delta\nu}\right)^{1/4}, \qquad (16)$$

i.e., at least of the order ten. But this number must not exceed $\nu/\Delta\nu$; hence we obtain the condition that, in order to realize optimum conditions, the frequency band must be so restricted that

$$\left(\frac{\nu}{\Delta\nu}\right) > 20{\cdot}2\left(\frac{c}{v}\right)^{2/3}. \qquad (17)$$

This is a somewhat surprising result. In the mathematical representation, it did not matter whether we divided up the frequency band into broad or narrow strips. But by the intervention of the number 137, $(20{\cdot}2 = (137/1{\cdot}5)^{2/3})$, it turns out that only narrow frequency bands are capable of accurate analysis by means of electrons. (If ions with charge Z were used, one would have to replace 137 by $137/\mathrm{Z}^2$, and the condition would be less stringent.)

Thus, at least in theory, the device could be perfected up to the optimum performance. The practical difficulties are of course evident. In practice one would rather replace the wave guide by a " high-Q " resonator, but this would somewhat complicate the theory.

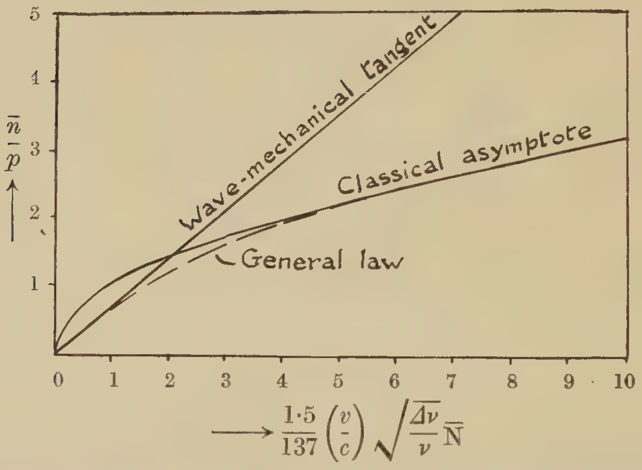
We thus find, assuming P passages, in the " classical " case

$$\left(\frac{\bar{n}}{\mathrm{P}}\right)^2 = \frac{1\cdot5}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}}\bar{\mathrm{N}}, \qquad (18)$$

while the corresponding wave-mechanical formula, valid for very weak interchange, is (Appendix III.)

$$\frac{\bar{n}}{\mathrm{P}} = \frac{1\cdot0}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}}\bar{\mathrm{N}}. \qquad (19)$$

Fig. 7.



Mean quantum exchange between electrons and photons. The initial tangent is calculated from wave mechanics, the asymptotic parabola classically.

Apart from the factor 2/3, the coefficient of $\bar{\mathrm{N}}$ is the same in both cases, but this time the quantum exchange $\bar{n}$ is proportional to the photon number itself, not to its square root It can be said, therefore, that *for small photon concentrations, the device acts as a counter, at large concentrations, as a field-measuring instrument.* The intermediate region is difficult to calculate, but, as shown in fig 7, the two branches can be connected by a plausible curve.

One can also interpret the results in this way: If there are few photons present, there will be few collisions, and—as shown in Appendix III.—equal probabilities of gains and losses, at any instant. With increasing photon concentration, repeated collisions will increase in number, and the resulting loss or gain increases with the square root of the photon number only, but this resultant has now a prevailing direction, which changes its sign with the frequency of the signal. At this stage the "classical field" has developed.

Having ascertained that, with certain reservations, we can make the exchange as strong as we like, we ask the question: If we know the average photon number $\overline{N}$, how must we adjust the electron beam current J, and the exchange parameter $\bar{n}$ in order to measure the field amplitude E with maximum accuracy? And having made these adjustments, how many steps shall we be able to distinguish in the scale of the field amplitudes? Evidently this question has a precise meaning only in the "classical" range of large $\bar{n}$ and $\overline{N}$, and the following considerations relate only to this case. In order to simplify the problem, we neglect the thermal noise, *i.e.*, we put $\overline{N}_T=0$, so that the relative accuracy on the photon scale would be $1/\overline{N}^{1/2}$, and the total number of steps in the photon ladder $2\overline{N}^{1/2}$. The calculations are carried out in Appendix IV., here we give only the physical considerations

The quantity to be measured is the electric amplitude in the information cell, which is proportional to the square root of the photon number. The measured quantity, on the other hand, is the alternating electron current, which, as mentioned above, is proportional to $\bar{n}.J$ or to $n.\overline{M}$ for not too strong signals, $\overline{M}$ being the mean number of exploring electrons per cell. For the optimum, we impose the condition that the relative mean square deviation of the quantity measured from the quantity to be measured shall be as small as possible, *i.e.*,

$$\frac{\overline{(nM-CN^{1/2})^2}}{(n\overline{M})^2}=\text{min}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

where the proportionality factor C is determined from the condition

$$\overline{(nM-CN^{1/2})}=0.$$

We have to choose $\bar{n}$ and $\overline{M}$ so as to satisfy the condition (20) for given $\overline{N}$. There must be an optimum, for these reasons: A too weak interchange $\bar{n}$ would leave the cell unexplored. A too strong interchange, on the other hand, will *interfere with the object of the measurement* and spoil it. Though in the mean electrons are as often accelerated as retarded, fluctuations in the numbers M_1 and M_2 of electrons which pass through accelerating and retarding phases might produce extra photons, which could not be distinguished from those belonging to the signal, or annihilate some. The spurious photons are generated according to a law

$$\delta_2N=\bar{n}(M_1-M_2)=\bar{n}(\delta M_1-\delta M_2) \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

as $\overline{M}_1=\overline{M}_2=\frac{1}{2}\overline{M}$. We have written $\bar{\delta}_2N$ for this number, considering it as a fluctuation which must be added to the natural fluctuation δ_1N, whose law is $\overline{\delta_1N^2}=\overline{N}$. The two fluctuations must be considered as independent.

It is already evident from the above that the fluctuations in the number of beam electrons, *i.e.*, the "shot effect", play an important part in these phenomena. A too weak current has a high relative fluctuation. A too strong current, especially aided by a large exchange, will again spoil the object. It may be noted that *we have here a type of uncertainty which springs directly from the fact that photons and electrons are discrete,* without any reference to the physical values of h and of e.*

The relative error according to equation (20) is calculated on the basis of equation (21), together with such evident assumptions as the independence of the fluctuations of n and M and of the "natural" part of δN. We assume also "normal" shot effect, $\delta M^2 = \bar{M}$. The result is

$$\frac{\overline{(nM - CN^{1/2})^2}}{(\overline{nM})^2} = \frac{1}{4\bar{N}} + \frac{1}{\bar{n}} + \bar{n}\left(\frac{1}{\overline{nM}} + \frac{\overline{nM}}{4\bar{N}^2}\right) . \quad . \quad . \quad . \qquad (22)$$

This is a minimum for

$$\overline{nM} = \bar{n} . \bar{M} = 2\bar{N} \quad \bar{n}^2 = \bar{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (23)$$

from which $\bar{M} = 2\bar{N}^{1/2}$. This gives the simple rule that, for optimum analysis of the signal, one must take one electron for every step in the scale of the photons, and the interchange $\bar{n}$ must be itself equal to one distinguishable step at the level $\bar{N}$. This again is a general rule, quite independent of the special model from which we started.

Substituting these values, we obtain for the minimum of the mean square relative error in the measurement of amplitudes

$$\left(\frac{\overline{(nM - CN^{1/2})^2}}{(\overline{nM})^2}\right)_{\min} = \frac{1}{4\bar{N}} + \frac{2}{\bar{n}} = \frac{1}{4\bar{N}} + \frac{2}{\bar{N}^{1/2}} . \quad . \quad . \quad . \qquad (24)$$

As we are dealing with large photon numbers only, the first term at the right-hand side can be neglected with respect to the second. Thus we see, applying the same rule which we have used in constructing the photon ladder, that the smallest distinguishable relative step in the amplitude scale is $\sqrt{2}$ times larger than the square root of the corresponding quantity in the photon scale. In other words, *the proper scale of the amplitudes will contain always less than the square root of the number of steps in the photon scale;* apart from the factor $\sqrt{2}$ for the reason that the optimum setting is, of course, possible for one level only.

Having determined the proper scale of the amplitudes, a simple application of the Uncertainty Principle (Appendix IV.) shows that the proper scale of the phase must also contain less steps than the square root of the photon scale. Thus, summing up, we see that the classical description of the signal, by being too detailed, gives in fact a somewhat *smaller total amount of information* than the quantum description.

The classical method of description, though theoretically inferior, may of course be still the best practically in the range of frequencies used for electrical communications, where efficient photon counters are not available. Conditions are different in the optical region, where detectors of the counter type—such as the eye—are not far from perfection. In this region, analysis in terms of electromagnetic waves is as yet technically impossible, but it is interesting to note that, even if it were possible, it would not be very practical. Progress in the field of microwaves is now actually approaching a region where the two different methods of analysis may become competitive. We see from our results that it takes about a hundred million photons per information cell in order to define amplitude and phase of the signal to 1 per cent each. Remembering that at 1 cm. wavelength the number of thermal photons per cell is only about 200, it may be seen that the time may not be far off when the imperfections of the classical method of description will manifest themselves even in electrical communications.

---

* It may be noted, however, that the shot effect, that is to say the random arrival of the exploring electrons, is a practical rather than a fundamental limitation. As Prof. R. E. Peierls has pointed out to me, it must be *ideally* possible to measure amplitude and phase simultaneously up to the limit given by the relation 12. In fact if one could make the exploring electrons arrive at regular intervals, all terms but the first on the right-hand side of equation 22 would vanish. But as it is practically impossible to suppress the shot effect to any significant degree, our results retain their practical importance.

Will it ever be possible to generate and to amplify light by an extension of microwave techniques ? Generation may not be unthinkable, but we can show at once that a light-amplifier for the reception of weak-light signals is impossible. It follows from our results that an electromagnetic signal in which amplitude and phase are defined with a relative accuracy ϵ must carry at least a power

$$h\nu^2\epsilon^{-4}(\Delta\nu/\nu). \quad \ldots\ldots\ldots \quad (25)$$

Applied to visible light with $\lambda=5000$ A., a line width of 1 A. and $\epsilon=0{\cdot}01$, we obtain 4·7 watts. Even if amplitude and phase were defined to 10 per cent only, the power would still be 0·0005 watts, a very high intensity in terms of visible light. This example demonstrates that *what we see* are always photons, not waves.

It may be hoped that these considerations have shown that the concepts of information theory may well prove their usefulness when applied to problems of physics.

REFERENCES.

BORN, M., 1949, *Natural Philosophy of Cause and Chance* (Oxford: University Press), p. 81.
CARSON, J. R., and FRY, T. C., 1937, *Bell Syst. T. J.*, **16,** 513.
GABOR, D., 1946, *Journ. I.E.E.*, **93,** III, 429 ; 1947, *Ibid.*, **94,** III, 369 ; 1947, *Nature, Lond.*, **159,** 591.
HEITLER, W., 1944, *The Quantum Theory of Radiation* (Oxford : University Press, 2nd edn.).
JORDAN, P., 1933, *Statistische Mechanik* (Braunschweig : Fr. Vieweg).
VON LAUE, M., 1914, *Ann. d. Phys.* [4] **44,** 1197.
MACDONALD, D. K. C., and KOMPFNER, R., 1949, *Proc. I.R.E.*, **37,** 1424.
MACKAY, D. M., 1950, *Phil. Mag.* [7] **41,** 189.
NYQUIST, H., 1928, *Phys. Rev.*, **32,** 753.
OSWALD, J., 1949, *C.R.*, **229,** 21.
OSWALD, J., 1950, *Cables et Transmission,* **2,** 197.
VAN DER POL, B., 1930, *Proc. I.R.E.*, **18,** 1194 ; 1946, *Journ. I.E.E.*, **93,** III, 153.
SHANNON, C. E., 1948, *Bell Syst. T.J.*, **27,** 379, 623 ; 1949, *Proc. I.R.E.*, **37,** 10
SMITH, L. P., 1946, *Phys. Rev.*, **69,** 195.
SZILÁRD, L., 1925, *Zeitschr. f. Phys.*, **32,** 753, 1925 ; 1929, *Ibid.*, **53,** 840.
TULLER, W. G., 1949, *Proc. I.R.E.*, **37,** 468.
VILLE, J. A., 1948, *Cables et Transmissions,* **1,** 61.

APPENDIX I.

ENERGY FLUCTUATIONS IN AN INFORMATION CELL IN THE PRESENCE OF A SIGNAL.

As in the case of thermal equilibrium, the mean square energy fluctuation is the sum of a classical and of a quantum term.

The classical term. Consider, for simplicity, a (complex) Fourrier-component E_S of the signal, and a component E_T of the thermal noise, corresponding to two different (circular) frequencies, say, ω_1 and ω_2. The instantaneous energy density, resulting from the interference, is proportional to

$$E_SE_S^*+E_TE_T^*+\{E_SE_T^*\exp[j(\omega_1-\omega_2)t]+E_S^*E_T\exp[-j(\omega_1-\omega_2)t]\},$$

where the asterisks stand for the conjugate complex values. The first two terms are the energy of the signal and the energy of the noise, the rest arises from interference. (Beats.) Let us write

$$\epsilon=\epsilon_S+\epsilon_T+\epsilon_{ST}.$$

The mean value of the interference energy ϵ_{ST} is *nil*, but its mean square is

$$\overline{\epsilon_{ST}^2}=2\overline{E_SE_S^*\,.\,E_TE_T^*}=2\bar{\epsilon}_S\epsilon_T.$$

Using this and the relations

$$\overline{\epsilon_S\epsilon_{ST}}=\overline{\epsilon_T\epsilon_{ST}}=0$$

which are evident, as there is no correlation between the signal and the noise, one obtains

$$\overline{(\epsilon-\bar{\epsilon})^2}=\overline{[(\epsilon_S+\epsilon_T+\epsilon_{ST})-(\epsilon_S+\bar{\epsilon}_T)]^2}=\overline{\epsilon_T^2}-\bar{\epsilon}_T^2+2\epsilon_S\bar{\epsilon}_T.$$

One knows the first two terms from Lorentz's calculation, which gives, in the absence of a signal

$$\overline{\epsilon_T^2}-\bar{\epsilon}_T^2=\bar{\epsilon}_T^2,$$

hence

$$\overline{(\epsilon-\bar{\epsilon})^2}=2\epsilon_S\bar{\epsilon}_T+\bar{\epsilon}_T^2,$$

or, as

$$\epsilon_S=\bar{\epsilon}-\bar{\epsilon}_T,$$

$$\overline{(\epsilon-\bar{\epsilon})^2_{\text{cl.}}}=(2\bar{\epsilon}-\bar{\epsilon}_T)\bar{\epsilon}_T. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (I.1)$$

This is the classical part of the fluctuation, arising from the interference of waves. *The quantum term* is always given by the " law of rare events "

$$\overline{(\epsilon-\bar{\epsilon})^2_{\text{qu.}}}=h\nu\bar{\epsilon} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (I.2)$$

as if the energy were present in the form of particles (photons), with energy $h\nu$. Adding (I.1) and (I.2), one obtains equation (4).

APPENDIX II.

CLASSICAL ENERGY EXCHANGE BETWEEN A WAVE AND AN ELECTRON BEAM.

In a rectangular wave guide with cross-section $a\times b$ (fig. 5), the components of the electric field are, in the TE_{01} mode,

$$E_x=E_0\cos\left(\frac{\pi y}{b}\right)\cdot\cos 2\pi\nu\left(t-\frac{z}{V}\right),\quad E_y=E_z=0,$$

where V is the phase velocity

$$V=c\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}.$$

Thus the mean electric energy is, per unit length of wave guide,

$$\frac{1}{8\pi}\frac{1}{4}E_0^2\,ab,$$

and the mean total energy, electric and magnetic, is twice as much. This energy moves with the group velocity $U=c^2/V$. The mean number $\overline{N}$ of photons in the information cell is obtained by equating the energy flux expressed by $\overline{N}$ and expressed by E_0. As the cell occupies a time $\varDelta t=1/\varDelta\nu$, the flux per second, *i.e.*, the power, is

$$h\nu\overline{N}\Delta\nu=\frac{1}{8\pi}\,\tfrac{1}{2}E_0^2abU=\frac{1}{16\pi}\,E_0^2abc\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}. \quad . \quad . \quad (\text{II}.1)$$

Consider an electron which traverses the guide on the x-axis. Assume that its velocity v is large enough to neglect its change during the transit. Thus the number of quanta exchanged—gained or lost—will be, in the mean,

$$\bar{n}=\frac{e}{h\nu}\overline{\left|\int_{-1/2a}^{1/2a}E\,dx\right|}=\frac{ev}{h\nu}\overline{\left|\int_{t_0}^{t_0+a/v}E\,dt\right|},$$

where the mean is to be taken with respect to the entrance time t_0. A simple calculation gives

$$\bar{n}=\frac{2}{\pi}\frac{\sin\theta}{\theta}\frac{ea}{h\nu}E_0 \qquad (\theta<\pi), \quad . \quad . \quad . \quad . \quad . \quad (\text{II}.2)$$

where θ is half the "transit angle," *i.e.*,

$$\theta=\pi\nu a/v.$$

The condition $\theta<\pi$ for the validity of (11.2) is always satisfied in an efficient arrangement, in fact, one must take $\theta<\frac{1}{2}\pi$.

Eliminating the field amplitude E_0 between (II.1) and (II.2), one obtains a direct relation between the photon number $\overline{N}$ and the exchange parameter $\bar{n}$:

$$\bar{n}^2=\frac{32}{\pi^2}\frac{\sin^2\theta}{\theta^2}\frac{a}{b}\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{-1/2}\frac{\varDelta\nu}{\nu}\overline{N}$$

$$=\frac{1}{137}\frac{32}{\pi^3}\frac{\sin^2\theta}{\theta}\frac{v}{\nu b}\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{-1/2}\frac{\varDelta\nu}{\nu}\overline{N}. \quad . \quad . \quad (\text{II}.3)$$

As explained in the text, the coefficient of $\overline{N}$ must be made as large as possible. The depth a appears only in the form $\sin^2\theta/\theta$, whose maximum value 0·723 is reached at $\theta=67°$, *i.e.*, at a transit angle of 134°. The width b, on the other hand, appears in the factor

$$1/b\,.\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}=c/bU.$$

The smallest admissible value of U is obtained if the group velocity is zero at the lower end of the frequency band, *i.e.*, if frequencies below $\nu=\frac{1}{2}\Delta\nu$ are cut off. In this case

$$b=\tfrac{1}{2}c/(\nu-\tfrac{1}{2}\Delta\nu),$$

hence the best value of the factor in question is the reciprocal of

$$b\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}=\frac{c}{2(\nu-\frac{1}{2}\Delta\nu)}\left[\frac{\Delta\nu}{\nu}\left(1-\frac{1}{4}\frac{\Delta\nu}{\nu}\right)\right]^{1/2}\approx\frac{c}{2\nu}\left(\frac{\Delta\nu}{\nu}\right)^{1/2},$$

assuming $\Delta\nu/\nu\ll 1$, *i.e.*, the band to be narrow. Substituting these values, one obtains

$$\bar{n}^2=\frac{1{\cdot}50}{137}\frac{v}{c}\left(\frac{\Delta\nu}{\nu}\right)^{1/2}\mathrm{N}, \qquad \text{(II.4)}$$

which is equation (15) of the text.

If repeated passages are used, it is evidently permissible to consider them of equal value, so long as their number P is small. It may be asked when this assumption will lead to an appreciable error. An estimate can be made by introducing the concept of the "instantaneous frequency"* which will vary slowly between the limits $\nu\pm\frac{1}{2}\Delta\nu$. A somewhat long calculation, which may be omitted, gives the result that for large P's the factor of $\bar{n}$ is not exactly P but approximately

$$\left(\overline{\frac{\sin \mathrm{P}\pi\delta\nu/\nu}{\pi\delta\nu/\nu}}\right)=\mathrm{P}\left[1-\frac{1}{6}\mathrm{P}^2\pi^2\left(\frac{\overline{\delta\nu^2}}{\nu}\right)+\dots\right],$$

where $\delta\nu$ is the deviation of the instantaneous from the mean frequency. Assuming that the instantaneous frequency is uniformly distributed in the available band $\Delta\nu$, one has $\overline{\delta\nu^2}=1/12(\Delta\nu)^2$. In the text, we have assumed that the maximum number of admissible passages is $\mathrm{P}_{\max}=\nu/\Delta\nu$. Even this gives an error of less than 14 per cent, which justifies our approximate treatment.

APPENDIX III.

ENERGY EXCHANGE OF ELECTRONS AND WEAK FIELDS ACCORDING TO QUANTUM MECHANICS.

We apply to the problem the standard perturbation method of wave mechanics. As the beam width can be considered as small in comparison to the spatial periods of the field, it is sufficient to start from Schrödinger's one-dimensional equation with vector potential A_x,

$$\frac{h}{2\pi i}\frac{\partial\Psi}{\partial t}=\frac{h^2}{8\pi^2 m}\frac{\partial^2\Psi}{\partial x^2}-\frac{ieh}{2\pi mc}A_x\frac{\partial\Psi}{\partial x}. \qquad \text{(III.1)}$$

We put

$$A_x=A_0\cos\omega_0 t, \qquad A_y=A_z=0, \qquad \omega_0=2\pi\nu_0,$$

---

* This concept is due to Helmholtz and has been first applied in communication theory by Balth. v.d. Pol. (1930). *Cf.* also J. R. Carson and T. C. Fry (1937).

which corresponds to an electric field

$$E_x = E_0 \sin \omega_0 t, \qquad E_y = E_z = 0, \qquad E_0 = A_0/\omega_0 c,$$

ν_0 is here the frequency of the field, which previously we have called ν, but in this calculation we will reserve ν for the frequency of the oscillators by which we will later replace the electron.

Assume that the solutions for zero field are written in the form

$$\Psi_n = \psi_n \exp\left(-i2\pi\epsilon_n t/h\right),$$

where ϵ_n is the energy of the unperturbed electron in some state " n ". The general solution of equation (III.1) can be written in the form

$$\Psi = \sum_m c_m(t)\Psi_m.$$

With this substitution, equation (III.1) becomes

$$\Sigma \frac{dc_m}{dt}\Psi_m = \frac{e}{mc}\Sigma c_m A_x \frac{\partial \Psi_m}{\partial x}. \quad . \quad . \quad . \quad . \quad . \quad \text{(III.2)}$$

Let us assume that the functions $\psi_m(x)$ are orthonormal

$$\int \psi_n^* \psi_m \, dx = \delta_{nm}.$$

Multiplying equation (III.2) by Ψ_r^*, integrating over the whole domain and making use of the orthonomality conditions, we obtain ordinary differential equations for the coefficients $c_r(t)$ in the form

$$\frac{dc_r}{dt} = i\frac{\pi e A_0}{mch}\sum_m c_m p_{mr}\{\exp\left[i(\omega_0+\omega_{rm})t\right] + \exp\left[-i(\omega_0-\omega_{rm})t\right]\}, \quad . \quad . \quad . \quad \text{(III.3)}$$

where
$$p_{mr} = \frac{h}{2\pi i}\int \psi_r^* \frac{\partial}{\partial x}\psi_m \, dx$$

is the matrix element of the momentum p of the electron, and

$$\omega_{rm} = \frac{2\pi}{h}(\epsilon_r - \epsilon_m)$$

is the circular frequency which corresponds to the transition $r \rightarrow m$.

Assume now that at the instant t_0 the electron is in the state $m=0$. Thus at the start the coefficient c_0 is unity, all the others are zero. Let us also assume that during the time of interaction, *i.e.*, during the passage of the electron, the other states c_m increase slowly enough to neglect transitions other than $0 \rightarrow r$. Thus we obtain, integrating equation (III.3) for the coefficient c_r at the time t,

$$c_r(t) = \frac{neA_0}{mch}p_{0r}\left\{\frac{\exp\left[i(\omega_0+\omega_r)t\right] - \exp\left[i(\omega_0+\omega_r)t_0\right]}{(\omega_0+\omega_r)} - \frac{\exp\left[-i(\omega_0-\omega_r)t\right] - \exp\left[-i(\omega_0-\omega_t)t_0\right]}{(\omega_0-\omega_r)}\right\}.$$

The absolute square of c_r is the probability that after the time $t-t_0$ the electron will be in the state " r ", having absorbed an energy corresponding to the frequency $\omega_{r0}=\omega_r$. Substituting E_0 instead of A_0, we find finally for the probability distribution of the electron at time t the expression

$$c_r c_r^* = \left(\frac{eE_0}{mhv_0}\right)^2 |p_{0r}|^2 \left\{\frac{\sin^2 \frac{1}{2}(\omega_0+\omega_r)(t-t_0)}{(\omega_0+\omega_r)^2} + \frac{\sin^2 \frac{1}{2}(\omega_0-\omega_r)(t-t_0)}{(\omega_0-\omega_r)^2} + 2\cos 2\omega_0 t \, \frac{\sin^2 \frac{1}{2}\omega_r(t-t_0)}{\omega_0^2-\omega_r^2}\right\} . \quad . \quad \text{(III.4)}$$

There are several interesting features in which this differs from the classical result. Note first, that the expression is even in ω_r, hence an electron has the same probability to gain or to lose a certain amount of energy. Note also that only the last term depends on the phase of the electric field at the instant t_0, the others depend only on the interaction time $t-t_0$. But even this last term has a frequency *twice* that of the electric field. It is rather doubtful whether this corresponds to an experimentally observable effect. Our calculation relates to weak fields, where $\bar{n}$ will be smaller than unity; but in order to observe even $\bar{n}=1$, the electron beam must be monochromatized to such an extent that its entrance phase becomes indefinite within a whole cycle. This is of no importance for our subsequent calculations, as we will consider only the mean value of (III.4), averaged over the entrance times t_0, in which the last term vanishes.

There remains now the problem to apply equation (III.4), which is a well-known result of wave mechanics, to our problem of an electron traversing the oscillating field between two conducting plates. This means calculating the matrix of the (mechanical) momentum p. We will evaluate this by a classical method, making use of the correspondence principle. This is justified in our case, as the de Broglie wavelength of the electrons is always very small with respect to the distance a between the plates.

In order to simplify the calculations, and to avoid singularities, we replace the electron by a plane density wave of total charge e (per unit cross-section)

$$\rho(x, t) = \frac{e}{\sqrt{(\pi)}v\tau} \exp\left[-(t-t_0-x/v)^2/\tau\right] . D(x). \quad . \quad . \quad \text{(III.5)}$$

$D(x)$ is an " annihilation factor " which is unity inside the guide, *i.e.*, for $-\frac{1}{2}a<x<\frac{1}{2}a$, and zero outside. Apart from this, *i.e.*, if the electron is at more than about $v\tau$ from the walls, the distribution is gaussian. τ is a small time, with which we will go ultimately to the limit zero. It will make the calculations easier if for the start we ignore the factor $D(x)$ and make use of it only towards the end.

We make use of the Fourier formula

$$\exp -[(t-x/v)/\tau]^2 = \sqrt{(\pi)}\tau \int_{-\infty}^{\infty} \exp\left[-(\pi\tau\nu)^2 - 2\pi i\nu(t-x/v)\right] d\nu.$$

This is to say that we decompose the electron into travelling waves of density

$$\rho_\nu \, d\nu = e/v \exp -(\pi\tau\nu)^2 \, . \, \exp\left[-2\pi i\nu(t - x/v)\right] d\nu.$$

We have put the entrance time $t_0=0$, as it plays no part in the mean values which we want to calculate. In the following formulæ we will drop also the factor $\exp -(\pi\tau\nu)^2$, which cuts off the spectrum at very high frequencies, but which is practically unity in the frequency region to be investigated.

The electric field produced by the electron follows from Poisson's equation

$$\frac{\partial \mathrm{E}}{\partial x} = 4\pi\rho.$$

Let us decompose this too into harmonic components by putting

$$\mathrm{E} = \int_{-\infty}^{\infty} \mathrm{E}_\nu \, d\nu.$$

One obtains
$$\mathrm{E}_\nu = i\frac{2e}{\nu} \exp\left[-2\pi i\nu\left(t - \frac{x}{v}\right)\right]$$

for the progressive field waves which accompany the *free* electron. But in a plane condenser with boundaries at $x=\pm\frac{1}{2}a$, this field induces surface charges

$$\sigma_\nu(\pm\tfrac{1}{2}a) = \mp\frac{1}{4\pi}\mathrm{E}_\nu(\pm\tfrac{1}{2}a) = \mp\frac{ie}{2\pi\nu}\exp\left[-2\pi i\nu\left(t \mp \frac{1}{2}\frac{a}{v}\right)\right],$$

which cut off the field beyond $x=\pm\frac{1}{2}a$.

We can now calculate the electric moment of the charges in the condenser, *i.e.*, of the electron charge and the surface charges, whose sum total is zero. The Fourier component of the moment in the frequency interval $\nu \, . \, d\nu$ is

$$d\nu\left\{\int_{-\frac{1}{2}a}^{\frac{1}{2}a} x\rho_\nu \, dx + \tfrac{1}{2}a[\sigma_\nu(\tfrac{1}{2}a) - \sigma_\nu(-\tfrac{1}{2}a)]\right\} = i\frac{\epsilon v}{2\pi^2\nu^2}\sin(\pi\nu a/v) \, . \, \exp(-2\pi i\nu t) \, d\nu.$$

We now equate this to the " standard oscillator " of electromagnetic theory, *i.e.*, to an electron of charge e which oscillates with a frequency ν around an equal and opposite charge, with an instantaneous amplitude X_ν. This gives

$$\mathrm{X}_\nu d\nu = i\frac{v}{2\pi^2\nu^2}\sin(\pi\nu a/v) \, . \, \exp(-2\pi i\nu t) \, d\nu.$$

The mechanical momentum of this equivalent oscillator is $m\dot{\mathrm{X}}_\nu \, d\nu$. Thus the momentum is, for the frequency interval ν , $d\nu$,

$$\mathrm{P}_\nu \, d\nu = \frac{mv}{\pi\nu}\sin(\pi\nu a/v) \, . \, \exp(-2\pi i\nu t) \, d\nu. \quad . \quad . \quad . \quad (\mathrm{III.6})$$

We cannot introduce this directly into equation (III.4), because that equation relates to discrete, steady states. In that case one has only to take the absolute square of the matrix elements to form the quantities $|p_{0r}|^2$ which figure in equation (III.4). But here we have to do with continuous states, and we must also remember that the oscillators, according to equation (III.6), replace the electron during the finite transit time only. In order to obtain the equivalents of the "oscillator strengths" $|p_{0r}|^2$, we must now multiply the momentum amplitudes $\mathrm{P}_\nu\, d\nu$ with the conjugate amplitudes $\mathrm{P}^*_\mu\, d\mu$ which belong to another interval μ, $d\mu$; we must integrate over the whole domain of the μ, and finally average over the transit time a/v. It is to say that we must calculate, instead of $|p_{0r}|^2$,

$$|p_{0\nu}|^2\, d\nu = 2\frac{v}{a}\, d\nu \int_{-\frac{1}{2}a/v}^{\frac{1}{2}a/v} dt \int_{-\infty}^{\infty} \mathrm{P}_\nu \mathrm{P}^*_\mu\, d\mu. \quad . \quad . \quad . \quad \text{(III.7)}$$

We had to add the factor two, because in wave mechanics ν is considered as a positive quantity, while in the classical theory it runs from $-\infty$ to $+\infty$. Substituting from (III.6),

$$|p_{0\nu}|^2 = 2\left(\frac{mv}{\pi}\right)^2 \frac{v}{a} \frac{\sin \pi\nu a/v}{\nu} \int_{-\frac{1}{2}a/v}^{\frac{1}{2}a/v} dt \int_{-\infty}^{\infty} \frac{\sin \pi\mu a/v}{\mu} \exp\left[2\pi i(\mu-\nu)t\right] d\mu.$$

After integration with respect to t, the double integral becomes

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin \pi\mu a/v}{\mu} \sin \frac{\sin \pi(\mu-\nu)a/v}{\mu-\nu}\, d\mu.$$

With the substitutions $x=(2\mu/\nu)-1$ and $\theta=\pi\nu a/v$, this is transformed into

$$\frac{2}{\pi\nu}\int_{-\infty}^{\infty} \frac{\sin \frac{1}{2}\theta(1+x) \,.\, \sin \frac{1}{2}\theta(1-x)}{1-x^2}\, dx = \frac{2}{\pi\nu}\int_{0}^{\infty} \frac{\cos \theta x - \cos \theta}{1-x^2}\, dx = \frac{1}{\nu} \sin \theta,$$

thus finally

$$|p_{0\nu}|^2\, d\nu = 2\left(\frac{mv}{\pi}\right)^2 \frac{v}{a} \frac{\sin^2 \theta}{\nu^2}\, d\nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(III.8)}$$

This means simply that during the transit time a/v the coherence region of any frequency ν is $\nu \pm v/2a$, with a bandwidth v/a, which is the reciprocal of the transit time. A frequency ν cooperates with any frequency μ inside this band in producing the oscillator strength.

Substituting (III.8) instead of $|p_{0r}|^2$ into equation (III.4), where we drop the last term whose mean value is zero, and replacing $t-t_0$ by the transit time a/v, we obtain

$$|c_\nu|^2\, d\nu = \frac{1}{h}\left(\frac{2\pi e^2}{hc}\right) \frac{\mathrm{E}_0^2 acv}{4\pi^3\nu_0^2} \frac{d\nu}{\nu^2} \sin^2 (\pi\nu a/v) \left\{\frac{\sin^2 \pi(\nu_0+\nu)a/v}{[(\nu_0+\nu)a/v]^2} + \frac{\sin^2 \pi(\nu_0-\nu)a/v}{[(\nu_0-\nu)a/v]^2}\right\}$$
$$. \quad . \quad . \quad \text{(III.9)}$$

This is the probability for an electron to be, after the passage, in a state ν, $d\nu$, that is to say, to have absorbed the energy $h\nu$. Thus the energy spectrum of the originally monokinetic electrons after the passage through the field is of the form

$$\mathrm{S}(\theta) = \frac{\sin^2 \theta}{\theta^2}\left\{\frac{\sin^2 (\theta+\theta_0)}{(\theta+\theta_0)^2} + \frac{\sin^2 (\theta-\theta_0)}{(\theta-\theta_0)^2}\right\}, \quad . \quad . \quad \text{(III.10)}$$

where the frequencies (and energies) are expressed by the half transit-angles

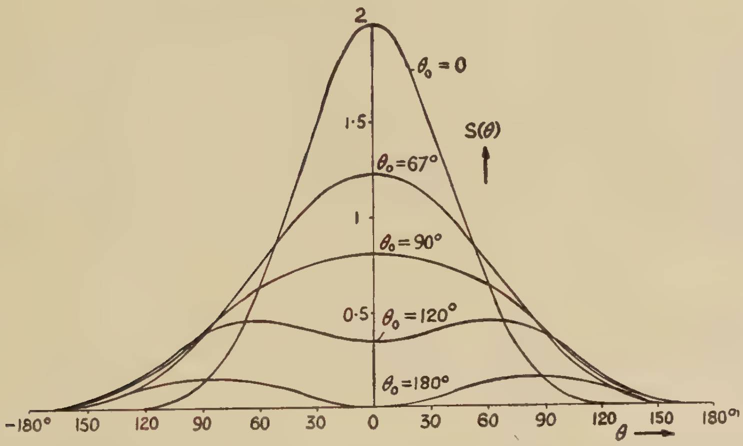
$$\theta=\pi\nu a/v, \qquad \theta_0=\pi\nu_0 a/v.$$

This function is shown in fig. 8 for a few values of θ_0. Experimental *checking*, though difficult, may not be impossible (*cf.* Macdonald and Kompfner 1949).

The mean energy exchanged with the field, lost or gained, is by (III.10)

$$\bar{n}h\nu_0=\int_0^\infty h\nu\,|c_\nu|^2\,d\nu=\frac{1}{137}\,\frac{E_0^2acv}{4\pi^3\nu_0^2}\,F(\theta_0), \qquad . \quad . \quad . \quad \text{(III.11)}$$

Fig. 8.



The function S(θ), equation (III.10), which gives the energy distribution of electrons after the transit. Gains and losses have equal probabilities. The "resonance point" at which the electron has gained or lost one quantum $h\nu_0$ of the oscillating field is at $\theta=\theta_0$.

where $F(\theta_0)$ is the integral

$$F(\theta_0)=\int_0^\infty \frac{d\theta}{\theta}\left[\frac{\sin^2(\theta+\theta_0)}{(\theta+\theta_0)^2}+\frac{\sin^2(\theta-\theta_0)}{(\theta-\theta_0)^2}\right].$$

For $\theta=67°$, which was the most favourable case in the classical theory, the value of the integral is found to be 1·22. Substituting this into equation (III.11) and expressing, as before the field E_0 by the number of photons, $\bar{N}$, by means of equation (II.1), and finally giving the wave guide the same width b which was found the best in the classical case, one finds

$$\bar{n}=\frac{1\cdot0}{137}\,\frac{v}{c}\left(\frac{\Delta\nu}{\nu}\right)^{1/2}\bar{N},$$

where we have again written ν for the field frequency, instead of ν_0. This is the energy exchange in weak fields, discussed in the text.

APPENDIX IV.

CONDITIONS FOR OPTIMUM EXPLORATION OF ELECTRIC FIELDS BY ELECTRONS.

As explained in the text, we adopt as the criterion of optimum measurement the condition that

$$\frac{\overline{(n\mathrm{M}-\mathrm{C}\mathrm{N}^{1/2})^2}}{(\overline{n\mathrm{M}})^2},$$

must be as small as possible. The proportionality factor C is given by the condition

$$\overline{(n\mathrm{M}-\mathrm{C}\mathrm{N}^{1/2})}=0,$$

or, as we are dealing here only with the "classical" case, in which the fluctuations are relatively small, $\mathrm{C}=\overline{n\mathrm{M}}\,.\,(\overline{\mathrm{N}})^{-1/2}$. As the fluctuations of n, the energy exchange parameter, and of M, the electron number, can be supposed as independent of one another, we can write $\overline{n\mathrm{M}}=\bar{n}\,.\,\overline{\mathrm{M}}$.

Let us now first calculate

$$\overline{n^2\mathrm{M}^2}=\overline{[(\bar{n}+\delta n)(\overline{\mathrm{M}}+\delta\mathrm{M})]^2}=\overline{(\bar{n}\overline{\mathrm{M}}+\bar{n}\delta\mathrm{M}+\overline{\mathrm{M}}\delta n)^2}=\bar{n}^2\overline{\mathrm{M}}^2+\bar{n}^2\overline{\delta\mathrm{M}^2}+\overline{\mathrm{M}}^2\overline{\delta n^2}$$

$$=\bar{n}^2\overline{\mathrm{M}}^2+\bar{n}^2\mathrm{M}+\overline{\mathrm{M}}^2\bar{n}.$$

Here we have neglected the flunctuation term of the highest order, and we have assumed

$$\overline{\delta n\delta\mathrm{M}}=0, \qquad \overline{\delta n^2}=\bar{n}, \qquad \overline{\delta\mathrm{M}^2}=\overline{\mathrm{M}}.$$

The first is the assumption of the relative independence of the fluctuations of n and M. The second is the assumption that the absorption of a quantum is independent of those previously absorbed or lost, which is certainly admissible so long as the electron velocity is not changed appreciably. The third assumption means shot effect without space charge smoothing. This is a logical assumption in a theory in which each electron is considered to interact singly with the field. With these formulæ, we can now calculate

$$\overline{(n\mathrm{M}-\mathrm{C}\mathrm{N}^{1/2})^2}=\overline{n^2\mathrm{M}^2}-2\mathrm{C}\overline{(\bar{n}+\delta n)(\overline{\mathrm{M}}+\delta\mathrm{M})(\overline{\mathrm{N}}^{1/2}+\tfrac{1}{2}\delta\mathrm{N}(\overline{\mathrm{N}})^{-1/2})}$$

$$+\overline{\mathrm{C}^2(\overline{\mathrm{N}}^{1/2}+\tfrac{1}{2}\delta\mathrm{N}(\overline{\mathrm{N}})^{-1/2})^2}=(\bar{n}\,.\,\overline{\mathrm{M}}-\mathrm{C}\overline{\mathrm{N}}^{1/2})^2+\bar{n}^2\overline{\mathrm{M}}+\overline{\mathrm{M}}^2\bar{n}+\tfrac{1}{4}\mathrm{C}^2\frac{\overline{\delta\mathrm{N}^2}}{\overline{\mathrm{N}}}, \quad . \quad . \quad . \quad (\mathrm{IV}.1)$$

where the first term is zero.

In order to calculate the last term, we consider the fluctuation $\delta\mathrm{N}$ of the photons as the sum of two independent components $\delta\mathrm{N}_1$ and $\delta\mathrm{N}_2$

$$\delta\mathrm{N}=\delta\mathrm{N}_1+\delta\mathrm{N}_2,$$

of which the first represents the "natural" fluctuations, subject to the law $\overline{\delta\mathrm{N}_1^2}=\overline{\mathrm{N}}$, and the second is due to the exploring electrons, *i.e.*, it is the excess of emitted over absorbed photons. Let M_1 be the number of beam electrons which pass through the field in a retarding phase, M_2 the number of accelerated electrons. As in the mean $\overline{M}_1=\overline{M}_2=\frac{1}{2}\overline{M}$, and as in the mean each electron produces or absorbs $\bar{n}$ photons, we can write

$$\delta_2 N=\bar{n}(M_1-M_2)=\bar{n}(\delta M_1-\delta M_2),$$

and

$$\overline{\delta_2 N^2}=\bar{n}^2(\overline{\delta M_1^2}+\overline{\delta M_2^2})=\bar{n}^2(\overline{M}_1+\overline{M}_2)=\bar{n}^2\overline{M},$$

where we have assumed $\overline{\delta M_1 \delta M_2}=0$, in accordance with our assumption of a normal, entirely random shot effect.

We now substitute these results into equation (IV.1), and replacing C^2 by its value $\bar{n}^2\overline{M}^2/\overline{N}$, we obtain for the mean square relative error in the measurement of the amplitudes

$$\frac{\overline{(nM-CN^{1,2})^2}}{n\overline{M}^2}=\frac{1}{4\overline{N}}+\frac{1}{\bar{n}}+\bar{n}\left(\frac{1}{\overline{nM}}+\frac{\overline{nM}}{4\overline{N}^2}\right). \quad . \quad . \quad \text{(IV.2)}$$

This is equation (24), discussed in the text.

It is not so easy to give an exact analysis of the errors in the measurement of phase, because devices for the running measurement of phase, with uniform error, are by no means simple. But the order of magnitude can be seen immediately if we apply the uncertainty principle to the exploring electrons, in the form

$$\delta\epsilon\delta t \approx 1.$$

δt is the uncertainty in the time at which the electron passes some fixed plane, *e.g.*, a fluorescent screen on which the waveform is recorded. The uncertainty in the energy ϵ is, assuming that the electrons were monokinetic at the start,

$$\delta\epsilon=h\nu\delta n,$$

n being the number of quanta which it may have absorbed from or lost to the field. But as δn is of the order $\bar{n}^{1/2}$, we obtain

$$\nu\delta t \approx (\bar{n})^{-1/2}.$$

In words, the oscillogram of the signal is traced with a spot whose mean width in the direction of time is a fraction $1/\sqrt{\bar{n}}$ of a cycle. Thus one can determine the phase with the same order of relative error as the amplitude.

CIV. *Density Variations in Aluminium.*

By J. L. SNOEK,
Philips Research Laboratories,
Eindhoven*.

[Received August 25, 1950.]

SUMMARY.

From careful determinations of the density of samples of aluminium in various conditions the conclusion is drawn that grain boundaries in aluminium may have a more spacious structure than is generally assumed.

COLD WORK AND DENSITY.

A priori considerations lead one to assume that the density of any metal, particularly however, that of metals with a close packed lattice structure will diminish as a result of cold work.

Various authors have found such a decrease (Burgers), but very careful work by Smart, Smith and Phillips on copper has shown that the decrease observed becomes less and less as the amount of oxygen in the lattice diminishes so that in all probability in extremely pure copper there is only a very small effect of cold work on the density ($<0{\cdot}01$ per cent).

The above serves to show that one has to be careful in drawing conclusions from results obtained on so-called pure metals. The only satisfactory procedure seems to be to vary the amount of a certain impurity and extrapolate the observed data towards concentration zero. In the majority of cases this has not been done and the results obtained remain open to serious doubt.

DENSITY AND NUMBER OF VACANCIES.

Our measurements on the density of aluminium in various conditions were started with the object of obtaining some direct data on the concentration of vacancies in this metal, the underlying idea being that in thermodynamical equilibrium this concentration would depend very much on the temperature and that it would be possible to " freeze in " a certain number of vacancies by cooling rapidly from a high temperature, or, if the process of annihilation were very slow to eliminate them by a process of cold work and annealing.

Aluminium as an object for research was preferred to gold and silver or copper on account of its light weight, which makes the measurement of the density roughly four times more accurate than is possible for the heavier elements. Magnesium, which is even more favourable in that respect, was discarded because of its non-cubic structure and the difficulty in cold-working it satisfactorily.

---

* Communicated by the Author.

Before describing the results, which were very unexpected, we shall give some details on the technique of the measurement and on the chemical composition of the samples.

TECHNIQUE OF THE MEASUREMENT.

Determination of the density on aluminium foil of half a millimeter thickness showed us—as was mentioned to us by Dr. W. Ch. van Geel of this laboratory before—that the surface absorbs variable amounts of water vapour. It is desirable, therefore, to make the surface play an unimportant part in the determination.

The analytical balance used—a new one, capable of an accuracy of 0·1 milligram—allowed of a maximum weight of 200 gram. As the standard shape of the object of which the density had to be determined we choose accordingly a cylinder with a diameter of 19 mm. and a length of 85 mm., having at one end a butt carrying a hole through which a molybdenum wire of 50 micron thickness was passed, serving to suspend the sample on one side of the balance.

The determination of the "dry" weight of the sample to within 0·1 milligram presented no undue difficulties, the weights used having been calibrated and the balance being placed in a room kept at constant temperature.

The determination of the weight in water presents more difficulties. Doubly distilled water was placed in a massive block made of aluminium containing four channels, two vertical ones stopping above the bottom of the block and two horizontal ones connecting the vertical holes. The vertical holes had diameters of 50 mm. subs 20 mm. each.

The wider vertical hole contained the sample, the smaller hole a Beckmann thermometer allowing the temperature of the water to be determined to within a few thousandths of a degree centigrade. The temperature never deviated more than 2 degrees from the temperature to which all density determinations were reduced, viz. 20° C. The usual corrections were of course applied.

The sample as a rule was turned on the lathe only a few hours before a determination was made, but measurements taken some days later showed no systematic deviations from this first value. Sharp corners were always avoided and the surface turned to a maximum of smoothness. The main difficulty lies in the capillary depression around the molybdenum wire. *Oleum pedis bovis* was finally chosen as a suitable means of reducing the uncertainties in the capillary behaviour to a minimum.

The question whether a given determination is satisfactory from this point of view is answered by a determination of the apparent sensitivity of the balance under the conditions of the experiment. The adding or taking off of one milligram, should always cause the same displacement of the needle. If it does not, something is wrong with the wetting of the molybdenum wire.

Heat-treatments as a rule were carried out on pieces of aluminium bar diameter 20 mm. as taken from stock. After the treatment the sample was given a suitable shape for a determination of the density. Such determinations are called a-measurements. For comparison an additional heat-treatment was sometimes applied to a cylinder which had already served for a measurement of the density. On account of the oxide-layer formed it was then first turned off on the lathe a second time. Measurements made on such a smaller specimen are called b-measurements.

IMPURITIES IN THE ALUMINIUM USED.

Aluminium was obtained in the form of long bars having a diameter of 20 mm. and a Vickers hardness of 29·2. Chemical analysis revealed the presence of the following impurity contents :

Impurities in aluminium used

Element	Per cent by weight
Si	0·006
Cu	0·003
Ga	0·002
Fe	0·004

The main impurity, however, is Mg, the amount of which was estimated from spectrochemical data to be at least 0·025 per cent by weight. No decrease in the Mg-content was observed after heat-treating the aluminium for many hours at 600° C.

TABLE I.

Number of specimen	Heat-treatment (in air)	$V_{\alpha ph}$	ρ_{20}		Mean grain-size (mm.)
1 (a)	As received	29·2	2·69885	85	—
2 (a)	,, ,,		85		
3 (a)	50 h. 400° C.	13·7	57	58	—
4 (a)	,, ,,		59		
5 (a)	,, ,, +2 h. 600° C.	14·4	81	84	—
6 (a)	,, ,, ,,		86		
7 (a)	,, ,, +20 h. 600° C.	13·3	78	86	—
8 (a)	,, ,, ,,		95		
9 (a)	,, ,, +200 h. 600° C.	13·6	89	92	—
10 (a)	,, ,, ,,		96		
1 (b)	50 h. 400° C.		40	42	0·5
2 (b)	,, ,,		43		
11 (a)	,, ,,		60	55	0·5
12 (a)	,, ,,		50		
9 (b)	quenched from 600° C. in water		90	95	6
10 (b)	quenched from 600° C. in water		100		
3 (b)	50 h. 400° C.+200 h. 600° C.		78	78	4
4 (b)	,, ,, ,, ,,		78		
5 (b)	,, ,, ,, ,,		92	94	4
6 (b)	,, ,, ,, ,,		95		
7 (b)	,, ,, ,, ,,		101	96	5
8 (b)	,, ,, ,, ,,		90		

EXPERIMENTAL RESULTS.

The outcome of our experiments was at first confusing and at any rate not in accordance with what we expected to find. It was found that heating at increasing temperatures caused the density to increase (Table I.) whereas quenching on the whole had no effect upon the measured density. The idea occurred to us that this might be an effect of the grain size on the mean density.

A second series of experiments was then started in which the temperature of the heating was kept constant at 600° C., the times of heating being varied between wide limits This led to essentially the same result (Table II.).

TABLE II.

Number of specimen	Heat-treatment	ρ_{20}	Mean grain-size (mm.)
13 (a)	2 min. 600° C. (salt bath)	2·69841	<1
14 (a)	5	76	1
15 (a)	10	68	2
16 (a)	25	76	2
17 (a)	50	80	3
18 (a)	100	81	3

TABLE III *.

(Fine grained as cast.)

Specimen	Heat-treatment	ρ_{20}		Mean grain-size (mm.)
A	As received	2·69975	$\Delta=+0{\cdot}00189$	<1
A	2 h. 600° C.	2·70164		6
B	As received	2·69981	$\Delta=+0{\cdot}00045$	<1
B	2 h. 600° C.	2·70026		4

As a third approach we had a cast made from a fairly cold melt, chilling it rapidly so that a fine grained structure was obtained. This material also became very coarse after a prolonged heating at 600° C. and by this treatment again became denser (Table III.). The aluminium used for this experiment came from another stock and contained less magnesium (0·005 per cent).

DISCUSSION OF RESULTS.

The above-mentioned results are hard to explain. One may first of all think of a spurious effect caused by impurities present in the material. The pursuit of this thought leads, however, to no tangible results, in view of

---

* The circumstance that the density of the cast samples on the whole is somewhat larger than for the extended bars may be due to the fact that it contained less magnesium.

the three very different series of heat-treatments applied to the aluminium, which all lead to essentially the same result. The cast sample besides contained considerably less magnesium than the wrought samples.

A second possibility is the appearance and gradual disappearance again of mysterious holes or gaps. We did not find a trace of such gaps on micro-etched samples.

Somewhat reluctantly we arrive therefore at the conclusion that grain boundaries in aluminium have a much more spacious structure than is generally supposed. They might, for instance, act as "retainers" for a number of vacancies which are formed at high temperature. No specific picture can be offered at the present time.

It is hoped that our experiments will be repeated by other physicists with a view to obtaining an independent check of the result presented here.

ADDENDUM.

These results have been discussed with Professor N. F. Mott and Dr. F. C. Frank who point out that if the density changes are attributed to an open structure at the grain boundary they imply that each boundary is equivalent to a void layer $\frac{1}{3}(5\times10^{-2}\times2\times10^{-4})$ cm$=330$ Å thick, which is hard to believe. It is almost as difficult to attribute the changes to precipitation of solutes, for if the whole of the analytically ascertained impurity were going into and coming out of solution, it would have to be changing volume by about 50 per cent. They have suggested that the aluminium specimens possibly contained gas (most probably hydrogen) which goes into solution at 600° C., stays in under ordinary cooling and particularly with water quench, but comes out again within 50 hours at 400° C. in the form of gas bubbles which lower the density.

This explanation cannot be entirely excluded, especially with regard to those samples (the majority) which came from a stock not prepared in the laboratory. The third series of measurements were carried out on a cast which had been made in the laboratory under the exclusion of free hydrogen, and in the presence of only very little water vapour. It seems unlikely that this cast would contain hydrogen in appreciable quantity. No holes were seen in microscopic examination of specimens after electropolishing, though one would expect to see them if gas bubbles were present to the extent required (one bubble of 30μ, 100 of 3μ or 10^6 of $0{\cdot}3\mu$ diamenter per grain).

Whatever the cause, the inherent difficulties connected with experiments of this kind on aluminium, and the desirability of an independent check, are emphasized.

REFERENCES.

BURGERS, W. G., *Handbuch der Metallphysik*, III **2**, 101.

SMART, J. S., SMITH, A. A., and PHILLIPS, A. J., 1941, *Amer. Inst. Min. Met. Eng.*, **143**, 272.

CV. CORRESPONDENCE.

Radiative Corrections in the Non-Relativistic Limit.

By W. Thirring, D.Phil.,
Dublin Institute for Advanced Studies *.

[Received September 10, 1950.]

In calculating radiative corrections to the Klein–Nishina formula it has been assumed by Corinaldesi and Jost (1948) and by Heitler and Ma (1949) that the corrections vanish in the non-relativistic limit. All terms proportional to the second-order matrix element were interpreted as renormalization, which means defining the electron-charge by the Thomson formula. The question is now, whether Dyson's (1949) renormalization procedure, the only one of which the consistency has been proved, leads to the same result, or whether there are finite corrections proportional to the Thomson cross-section analogous to the case of the magnetic moment. It has been shown by Schafroth (1949), by direct calculation, that after renormalization the fourth-order matrix element vanishes for zero momentum of the incident photon. We will give a simple proof that this holds in all approximations.

One obtains the graph of the nth order to the Compton effect by inserting successively the two external photon lines in any of the electron lines of the electron self-energy graphs of order $n-2$. An electron self-energy graph consists of the main electron line and some closed electron lines and internal photon lines. If one inserts the two external photon lines into the main electron line (energy-momentum $p\mu$) for the limit of zero momentum of the photon, using

$$\frac{\partial}{\partial p_\mu}(\mathbf{p}-m)^{-1}=-(\mathbf{p}-m)^{-1}\gamma_\mu(\mathbf{p}-m)^{-1}\,;\quad \mathbf{p}=p_\mu\gamma^\mu \qquad . \quad . \quad (1)$$

and summing over all insertions and all self-energy graphs, one obtains

$$(\mathbf{p}-m)\left[e^1_\mu e^2_\nu\frac{\partial}{\partial p_\mu}\frac{\partial}{\partial p_\nu}\mathrm{S}'(p)\right](\mathbf{p}-m) \qquad . \quad . \quad . \quad . \quad . \quad (2)$$

for the matrix element. After renormalization S′(p), as defined by Dyson, has the form

$$\mathrm{S}'(p)=(\mathbf{p}-m)^{-1}[1+\mathrm{B}(p^2-m^2)]+m^{-1}\mathrm{C}(p^2-m^2)\,\dagger, \qquad . \quad . \quad (3)$$

---

* Communicated by the Author.

† Dyson's renormalization casts contributions from proper self-energy parts into the form (3) whilst on multiplying the proper self-energy parts the contribution from the general self-energy part will be

$$(\mathbf{p}-m)^{-1}[1-\mathrm{B}-\mathrm{C}m^{-1}(\mathbf{p}-m)]^{-1}$$

but one readily sees that this is again of the form (3).

where B(0)=0. Inserting into (2) we get

$$(2)=[\mathbf{e}^1(\mathbf{p}-m)^{-1}\mathbf{e}^2+\mathbf{e}^2(\mathbf{p}-m)^{-1}](1+\mathrm{B}(0))-2[\mathbf{e}^1(pe^2)+e^2(pe^1)]\mathrm{B}'(0)$$
$$+4(e^1p)(e^2p)[(\mathbf{p}-m)\mathrm{B}''(0)+(\mathbf{p}-m)^2m^{-1}\mathrm{C}''(0)].$$

Everything but the uncorrected matrix element vanishes since one can choose for the potentials of the external photons a gauge system in which $(pe^1)=(pe^2)=0$.

Contributions from the closed loops will have the form

$$\int dq\, tr \prod_i (\mathbf{q}+\mathbf{k}_1-m)^{-1}e_i,$$

where k_i are some energy-momentum fourvectors, over some of which one might have to integrate. Inserting one external photon line successively in all electron lines one gets for zero momentum

$$-e_\mu \int dq \frac{\partial}{\partial q_\mu}\, tr \prod_i (\mathbf{q}+\mathbf{k}_i-m)^{-1}e_i,$$

which vanishes on partial integration, provided that the integral converges. Otherwise one has to apply regularization in order to avoid contributions from the boundary.

It is evident that the same argument applies to a case where the electron interacts any number of times with the electromagnetic field, and that therefore in all those cases the radiative corrections vanish in the non-relativistic limit.

References.

Corinaldesi, E., and Jost, R., 1948, *Helv. Phys. Acta*, **21**, 183.
Dyson, F. J., 1949, *Phys. Rev.*, **75**, 1736.
Heitler, W., and Ma, S. T., 1949, *Phil. Mag.*, **40**, 651.
Schafroth, M. R., 1949, *Helv. Phys. Acta*, **22**, 501.

Multiple Processes in the Production of Pairs of Electrons by γ-Radiation.

By J. E. Hooper and D. T. King,
H. H. Wills Physical Laboratory*.

[Received September 28, 1950.]

[Plate XXXV.]

It is shown by Heitler (1944) that the cross-section for the simultaneous emission of n quanta by a multiple Bremsstrahlung process is, at high energies, of the order $Z^2r_0^2/137^n$. On this basis one would expect the cross-section for the analogous case of the production of n pairs of electrons by a γ-ray in the electrostatic field of a nucleus to be of the same order of magnitude.

---

* Communicated by Professor C. F. Powell.

We have recently made an investigation of the pairs of electrons created in photographic emulsions exposed at an altitude of 68,000 ft., and resulting from the passage of γ-radiation arising in the decay of π^0-particles. Among 1380 pairs we have found two examples which appear to correspond to the simultaneous production of two pairs of electrons. A mosaic of photo-micrographs of one of these events is shown in Pl. XXXV. In both cases the associated tracks are of sufficient length to allow us to determine the grain-density in the tracks and the momenta of the particles by the scattering method. All the particles have a specific ionization equal to the minimum value for a particle of charge e, and the total energies of the particles in the two events are 220 and 2000 MeV. respectively.

The energies of the individual particles are shown in the Table.

	Event (A)	Event (B)
E_{1-4} (MeV.)	40 ± 5, 95 ± 10, 60 ± 8, 22 ± 4	8 ± 3, 960 ± 150, 20 ± 4, 980 ± 150
E_T (MeV.)	220 ± 25	2000 ± 300

These measurements show that nearly all the particles are less massive than protons. Further, in one case two and in the other, three of the particles are less massive than μ-mesons. These facts, and the observed even number of secondary particles, allow us to exclude the possibility that either of the events is due to the collision of a fast neutron with a proton in the emulsion in which a shower of π-mesons is created. We therefore attribute them to the production of pairs of electrons and positrons.

The possibility remains that the events are due to a process first observed by Bradt and Peters (1950) and by Occhialini (1949), in which one of the electrons of a pair, created in the emulsion by a γ-ray, subsequently generates a second pair in the field of a nucleus. We have observed 52 examples of pair creation by a fast particle—the so-called “ tridents ”—and it would be possible to explain the present observations in terms of a similar occurrence in which the production of the second pair has occurred, by chance, very close to the point of origin of the “ primary ” pair. Such an event would be difficult to distinguish from one involving the simultaneous production of four electrons in the field of a nucleus, especially in the case of pairs of small angular divergence such as are commonly produced by γ-rays of great quantum energy. We believe, however, that the following argument allows us to exclude this possibility :—

The number of observed events in which a “ trident ” is found to be associated with a pair of electrons enables us to estimate the average “ mean free path ” in the emulsion for trident production. The mean value is about 100 cm. The probability of the trident occurring within 10 μ of the point of origin of the pair is therefore 1 in 10^4; and the probability of observing two such events among 1300 pairs is about 1 in 10^5. These values are so small that we can disregard this explanation of the origin of the two events under discussion. Further, the probability that, by chance, two unrelated pairs should be superposed is less than 1 in 10^{12} per cm.^{2} of the emulsion in the conditions of the experiment. We therefore regard our observations as providing strong evidence for the multiple production of pairs of electrons by a γ-ray quantum in the field of a nucleus, the probability of the process being of the order of magnitude suggested by quantum-electrodynamics.

The observations on the " tridents " associated with pairs of electrons will enable us to determine the variation with energy of the cross-section for the production of " pairs " by fast electrons. Preliminary results suggest that there is a rapid increase in the value of the cross-section for the process as the energy of the incident electron increases. We may direct attention to the fact that some caution should be observed in observations of this type because of the possibility of confusion due to "Bremstrahlung". A fast electron loses energy in inelastic collisions which result in the creation of quanta. Especially in the case of electrons of high energy, these quanta may be emitted in directions nearly coincident with the line of motion of the recoiling electron. If, further, such a quantum produces a " pair ", the resulting electrons may appear to arise at a point nearly coincident with the track of the original electron. Such events can be distinguished only with difficulty from true trident production. The possibility of this source of confusion was first pointed out to us by Professor Peierls at the Edinburgh Conference in 1949.

REFERENCES.

HEITLER, 1944, *Quantum Theory of Radiation*, 2nd Ed., p. 181 (Oxford University Press).

BRADT and PETERS, 1950, *H.P.A.*, **24**, 23.

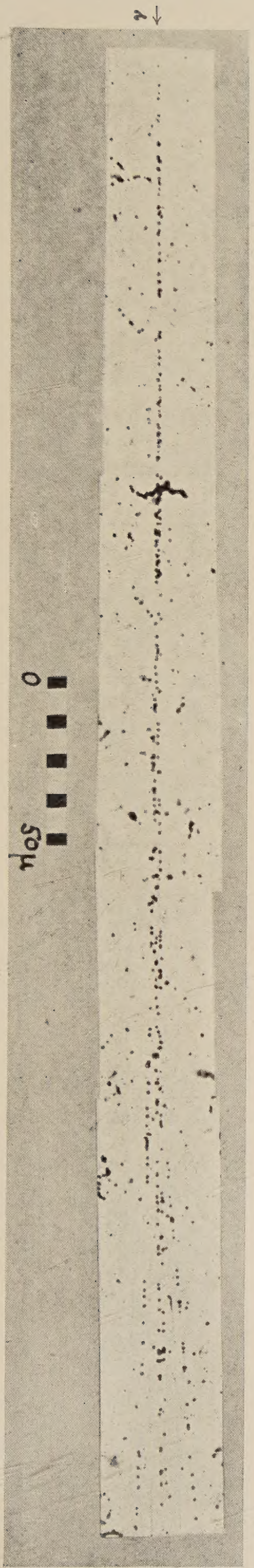
OCCHIALINI, 1949, *Como Conference Reports.*

CVI. *Notices of New Books and Periodicals received.*

Practical Applications of Spectrum Analysis. By H. DINGLE. [Pp. 245.] 14 plates. (Chapman and Hall, 1950.) Price 40*s*. net.

IN a general conclusion to this useful book the author states that in his experience more than half of the spectrographic analyses which have been published during the last thirty years are untrustworthy. Used in conjunction with a standard set of wave-length tables, this book should do much to remedy this state of affairs.

No attempt is made to describe the theoretical laws of spectroscopy, but the relevant aspects of the mechanism of radiation, of sources of luminosity and of the optical principles of the spectrograph are dealt with in so constructive a fashion as to lead to a confident and critical choice and use of the most appropriate apparatus for a very wide range of spectrographic analyses.



A mosaic of photo-micrographs illustrating an event in which two pairs of electrons appear to be created simultaneously by a 220 MeV. γ-ray.

The chapters on the measurement and interpretation of the spectrograms obtained in these analyses, together with the comprehensive tables of Raies Ultimes and the wealth of critical comment on special points of interpretation, taken element by element, embody great experience and form a valuable guide to the less experienced and an indispensable reference work to the more experienced workers in this field.

The writing throughout is clear in style, and the general level of production excellent, except for the reproduction of the plates, the quality of which is somewhat disappointing for a book at this price. D. McG.

Progress in Coal Science. Edited by D. H. BANGHAM. [Pp. 456.] (Butterworths Scientific Publications, London, 1950.)

THIS book comprises twenty-six review articles, many of which have appeared in the Monthly Bulletin of the British Coal Utilisation Research Association during the years 1943–1947. They are divided into the following sections :—Modern Experimental Techniques, Fine Particles, Constituents of Coal, Organic Chemistry of Coal Products, and Chemical Aspects of Combustion and Gasification.

The first section contains articles on crystallography, absorption spectrography, mass spectrometry, and chromatography, all as applied to chemical analysis. These should be of use to chemists generally, inside and outside the Coal industry, as also should the articles on the production and measurement of fine particles. The third section contains among others an account by Prof. V. M. Goldschmidt of his work on the rare elements in coal ash, and a review by Dr. C. A. Seyler of his work on coal petrology. In the fourth section is a translation of a German article by O. Gruber on " The Progress in Coal Tar Research in the period 1910 to 1940," and a review of the Fischer-Tropsch reaction. The last section contains an article by Prof. D. T. A. Townsend, which gives a general account of producer gas research, and more specialised articles on fuel beds, combustion kinetics, and other topics.

The book should be most valuable to all scientists in the coal and allied industries and, as indicated, some of the articles should find a wide appeal.

D. D. E.

Astronomy of Stellar Energy and Decay. By MARTIN JOHNSON. [Pp. 216, Plates 4.] (Faber and Faber, 1950.) Price 18*s*. 0*d*.

THIS book, which aims at bridging the gap between popular astronomy and the research literature of stellar physics, can be strongly recommended. It is deliberately set in two parts. The first, though framed for the general intelligent reader, can also be read with interest by students of wider scientific training. Part II. is labelled " student's introduction to analysis of the facts and argument of the theories." As its title suggests it is more qualitative, though never heavily mathematical, in its treatment. It supplements Part I. and serves as a very useful introduction to more advanced treatises. The two parts together can be specially recommended for private reading at the university degree stage. A. M. T.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]